

OM of: US-09-049-696-7 to: SPTREMBL_17.* out_format : pfs

Date: Mar 30, 2002 2:46 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+ntp.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09049696/runat_28032002_145238_2053/app_query.fasta_1.12579
-DB=SPTREMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1_705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-7
Query length: 289
Database: SPTREMBL_17.*
Database sequences: 473505
Database length: 146272329
Search time (sec): 805.760000

score_list:

Sequence	Strd Orig	ZScore	RScore	Len	Documentation
sp_human:Q9UNF7	+ 315.00	634.89	4.7e-28	917	Q9unf7 homo sapiens (human). ca
sp_rodent:Q9R070	+ 289.00	580.89	4.9e-25	902	Q9r070 mus musculus (mouse). ca
sp_rodent:Q9EQR4	+ 289.00	580.89	4.9e-25	902	Q9eqr4 mus musculus (mouse). en
sp_rodent:Q88860	+ 286.00	574.65	1.1e-24	901	Q88860 mus musculus (mouse). ch
sp_rodent:Q9QX15	+ 286.00	574.65	1.1e-24	902	Q9qx15 mus musculus (mouse). ca
sp_mammal:O18742	+ 282.00	567.31	3.2e-24	794	Q18742 bos taurus (bovine). lu
sp_mammal:O18743	+ 282.00	567.06	3.2e-24	820	Q18743 bos taurus (bovine). lu
sp_mammal:O18741	+ 282.00	566.29	3.2e-24	905	Q18741 bos taurus (bovine). lu
sp_human:Q9QVC9	+ 218.00	432.72	8.4e-17	943	Q9qvc9 homo sapiens (human). ch
sp_human:Q9V6N2	+ 218.00	432.72	8.4e-17	943	Q9v6n2 homo sapiens (human). ca
sp_human:Q9NXP1	+ 169.00	336.14	4.0e-11	469	Q9npx1 homo sapiens (human). co
sp_invertebrate:Q21317	+ 80.00	155.47	0.8462	259	Q21317 caenorhabditis elegans
sp_invertebrate:Q18005	+ 79.50	149.23	0.9661	505	Q18005 caenorhabditis elegans
sp_bacteria:Q9ABQ3	+ 79.50	146.83	0.9657	687	Q9abq3 caulobacter crescentus
sp_fungi:O14436	+ 79.00	145.34	1.10	728	Q14436 aspergillus fumigatus (s
sp_invertebrate:O44525	+ 77.50	148.21	1.65	337	Q44525 caenorhabditis elegans
sp_invertebrate:O01328	+ 77.50	145.07	1.65	505	Q01328 caenorhabditis elegans
sp_archaea:Q58581	+ 76.50	143.88	2.15	450	Q58581 pyrococcus horikoshii. 4
sp_invertebrate:O01334	+ 74.50	142.93	3.67	298	Q01334 caenorhabditis elegans
sp_human:Q15010	+ 73.50	125.67	4.79	2093	Q15010 homo sapiens (human). r
sp_plant:Q9S822	+ 73.00	131.09	5.48	912	Q9s822 zea mays (maize). heat s
sp_plant:Q9FN27	+ 72.50	137.53	6.27	349	Q9fn27 arabidopsis thaliana (mc
sp_invertebrate:O44523	+ 72.50	137.33	6.27	358	Q44523 caenorhabditis elegans
sp_invertebrate:Q918N5	+ 71.50	128.60	8.17	841	Q918n5 rana pipiens (northern l
sp_plant:Q9ZV10	+ 71.00	129.11	9.34	689	Q9zvl0 arabidopsis thaliana (mc
sp_invertebrate:O44524	+ 70.50	130.32	10.68	516	Q44524 caenorhabditis elegans
sp_virus:Q9PYX8	+ 70.00	129.85	12.21	480	Q9pyx8 xestia c-nigrum granulosa
sp_bacteria:Q9S4A4	+ 69.50	128.96	12.21	538	Q9s4a4 actinobacillus actinomyc
sp_plant:Q9ASW2	+ 69.50	134.04	13.97	245	Q9asw2 arabidopsis thaliana (mc
sp_plant:Q9SLK3	+ 69.50	133.58	13.97	260	Q9slk3 arabidopsis thaliana (mc
sp_bacteria:Q86590	+ 69.00	129.83	13.96	421	Q86590 streptomyces coelicolor.
sp_bacteria:Q9ZBN7	+ 69.00	128.46	15.95	439	Q9zbn7 wolinnella succinogenes.
sp_bacteria:Q57377	+ 68.50	134.49	18.25	177	Q57377 shigella flexneri, esche
sp_bacteria:Q9RIN0	+ 68.50	131.93	18.24	246	Q9rin0 streptococcus pneumoniae
sp_bacteria:O52560	+ 68.50	128.08	18.23	403	Q52560 amycolatopsis mediterr
sp_bacteria:O78827	+ 68.50	127.76	18.23	420	Q78827 ericoaulon aquaticum. fl
sp_invertebrate:Q213635	- 68.50	122.74	18.21	801	Q213635 caenorhabditis elegans
sp_plant:Q9XIA5	+ 68.00	128.19	20.84	348	Q9xia5 arabidopsis thaliana (mc
sp_fungi:O14239	+ 68.00	118.85	20.80	1155	Q14239 schizosaccharomyces pom

seq_name: sp_human:Q9UNF7

seq_documentation_block:

ID Q9UNF7 PRELIMINARY; PRT; 917 AA.
AC Q9UNF7;
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE CALCIIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 2.
GN CACCC2.

QC Homo sapiens (Human).
CS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RC SEQUENCE FROM N.A.

RP TISSUE=COLON;

RX MEDLINE=99364503; PubMed=10437792;

RA Agnel M., Vermet T., Culouscou J.M.;

RT Identification of three novel members of the calcium-dependent chloride channel (CACCC) family predominantly expressed in the digestive tract and trachea.;

RL FEBS Lett. 455:295-301(1999).

DR EMBL; AF127035; AAD48398.1; .

DR InterPro: IPR002035; WFA.

DR Pfam: PF00092; vwa; 1.

DR PROSITE; PS00234; WFA; 1.

DR SMART; SM00327; WFA; 1.

SQ SEQUENCE 917 AA; 101153 MW; EA01090D781BEB95 CRC64;

alignment_scores:

Quality: 315.00 Length: 92
Ratio: 3.889 Gaps: 0
Percent Similarity: 88.043 Percent Identity: 66.304

alignment_block:

us-09-049-696-7 x Q9UNF7

Align_seg 1/1 to: Q9UNF7 from: 1 to: 917

14 GATGATCTGAATGTGCTGCTCAGCGATGGGAAGACACACATATAG 63
|||||
404 ASpglySerGluValLeuLeuThraspglyGluaspasThralase 420
|||||
64 TGGTGTCTTAAAGAGTCAACAAAGTNGCCATCATCCACACAGTCG 113
|||||
420 rSerCysileAspgluVallyGlnSerGlyAlaileValHisphelLea 437
|||||
114 CTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAAATGACA 163
|||||
437 laLeuGlyArGlaAlaAspgluAlaValIleGluMetSerLysileThr 453
|||||
164 GGAGGTTCAGACATAGTCTCAGATCAAGTTCAGAACAAATGCCTCAT 213
|||||
454 GlycylSerHisPhetYrValSerAspgluAlaGlnAsnAsnGlyLeuIl 470
|||||
214 TGATGCTTTTGGGGCCCTTCATCAGGAAATGAGAGCTGTCTCTCAGCGCT 263
|||||
470 eAspAlaPhelGlyAlaLeuThrSerGlyAsnThraspLeuSerGlnlyss 487
|||||
264 CCATCCAGCTTGAGAGTANGGATTA 289
|||||
487 erLeuGlnLeuGluSerLysGlyLeu 495

seq_name: sp_rodent:Q9R070

seq_documentation_block:

ID Q9R070 PRELIMINARY; PRT; 902 AA.
AC Q9R070;
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)

Best Available Copy

DE CA(2+)-SENSITIVE CHLORIDE CHANNEL 2.
 GN CLCA2 OR CACC.
 OS Mus musculus (Mouse).
 RX Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=MAMMARY;
 RC MEDLINE=20012773; PubMed=10544033;
 RX Lee D., Ha S., Kho Y., Kim J., Cho K., Baik M., Choi Y.;
 RA "Induction of mouse Ca(2+)-sensitive chloride channel 2 gene during
 RT involution of mammary gland.";
 RL Biochem. Biophys. Res. Commun. 264:933-937(1999).
 DR EMBL; AF108501; AAF12731.1; -
 DR MGD; MGI:1931471; Clca2.
 DR InterPro; IPR002035; VWFA.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS50234; VWFA; 1.
 SQ SEQUENCE 902 AA; 99866 MW; 0FB746CF5FB3D07F CRC64;

alignment_scores:
 Quality: 289.00 Length: 91
 Ratio: 3.612 Gaps: 1
 Percent Similarity: 87.912 Percent Identity: 67.033

alignment_block:

US-09-049-696-7 x Q9R070 ..

Align seg 1/1 to: Q9R070 from: 1 to: 902

```

11 ACTGATGGATCTGAAATGTGCTGCTGACGAGTGGGAAGACACACTAT 60
|||||
405 ThrSerGlySerGluIleValLeuLeuThrAspGlyGluAsnGlyI1 421
61 AAGTGGTGCTTTAAGCAGGTCAACAAAGTNGTGCATCATCCACACAG 110
|||||
421 eSerSerCysPheGluAlaValSerArgSerGlyAlaIleIleHisThrI 438
111 TCGTTTGGGGCCCTCTGACGCTCAAGACTAGAGGAGTGTCCAAATG 160
|||||
438 leAlaLeuGlyProSerAlaAlaArgGluLeuGluThrLeuSerAspMet 454
161 ACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGAACAAATGGCCT 210
|||||
455 ThrGlyGlyLeuArgPheTyrAlaAsnLysHisVal.....SerSerLe 469
211 CATTCATGCTTTTGGGGCCCTTTCATCAGGAATGGAGCTGTCTCTCAGC 260
|||||
469 uileAspAlaPheSerArgIleSerThrSerGlySerValSerGlnG 486
261 GCTCCATCCAGCTTGAGAGTAAG 283
486 InAlaLeuGlnLeuGluSerLys 493

```

seq_name: sp_rodent:Q9EQR4

seq_documentation_block:

Q9EQR4 PRELIMINARY; PRT; 902 AA.

```

AC Q9EQR4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ENDOTHELIAL CHLORIDE CHANNEL.
GN CLCA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57/BL6;

```

RA Elble R.C., Pauli B.;
 RT "A murine endothelial homolog of Lu-ECAM-1 associated with lung
 RL melanoma metastasis.";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF115852; AAG47626.1; -
 DR InterPro; IPR003961; FN.III.
 DR InterPro; IPR002035; VWFA.
 DR SMART; SM00060; FN3; 1.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS50234; VWFA; 1.
 SQ SEQUENCE 902 AA; 99855 MW; F2BA57DAA5D4AAAD CRC64;

alignment_scores:
 Quality: 289.00 Length: 91
 Ratio: 3.612 Gaps: 1
 Percent Similarity: 87.912 Percent Identity: 67.033

alignment_block:

US-09-049-696-7 x Q9EQR4 ..

Align seg 1/1 to: Q9EQR4 from: 1 to: 902

```

11 ACTGATGGATCTGAAATGTGCTGCTGACGAGTGGGAAGACACACTAT 60
|||||
405 ThrSerGlySerGluIleValLeuLeuThrAspGlyGluAsnGlyI1 421
61 AAGTGGTGCTTTAAGCAGGTCAACAAAGTNGTGCATCATCCACACAG 110
|||||
421 eSerSerCysPheGluAlaValSerArgSerGlyAlaIleIleHisThrI 438
111 TCGTTTGGGGCCCTCTGACGCTCAAGACTAGAGGAGTGTCCAAATG 160
|||||
438 leAlaLeuGlyProSerAlaAlaArgGluLeuGluThrLeuSerAspMet 454
161 ACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGAACAAATGGCCT 210
|||||
455 ThrGlyGlyLeuArgPheTyrAlaAsnLysHisVal.....SerSerLe 469
211 CATTCATGCTTTTGGGGCCCTTTCATCAGGAATGGAGCTGTCTCTCAGC 260
|||||
469 uileAspAlaPheSerArgIleSerThrSerGlySerValSerGlnG 486
261 GCTCCATCCAGCTTGAGAGTAAG 283
486 InAlaLeuGlnLeuGluSerLys 493

```

seq_name: sp_rodent:Q88860

seq_documentation_block:

Q88860 PRELIMINARY; PRT; 901 AA.

```

AC Q88860;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CHLORIDE CHANNEL CACC.
GN CLCA1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RA Rocio L., Musante L., Cintil R., Moran O., Seri M., Galletta L.J.V.;
RT "Characterization of a murine gene homologous to the bovine CACC
  RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF052746; AAC35003.1; -
DR MGD; MGI:1316732; Clca1.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 901 AA; 100038 MW; 13DF068BF3337AAB CRC64;

```


alignment_scores:
 Quality: 286.00 Length: 91
 Ratio: 3.620 Gaps: 1
 Percent Similarity: 86.813 Percent Identity: 65.934

alignment_block:
 US-09-049-696-7 x 088860 ..
 Align seg 1/1 to: 088860 from: 1 to: 901

```

11 ACTGATGATCTGAATTTGCTGCTGACGATGGGGAAGACACACTAT 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
405 ThrSerGlySerGluIleValLeuThrAspGlyGluAspAsnGlyI 421
61 AAGTGGGTGCTTTAAACGAGGTCAACAAAGTNGTGCCATCATCCACACAG 110
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
421 eArgSerCysPheGluAlaValSerArgSerGlyAlaIleIleHisThrI 438
111 TCAGTTTGGGGCCCTTCGACGCTCAAGAACTAGAGGAGCTGTCCAAATG 160
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
438 leAlaLeuGlyProSerAlaAlaArgGluLeuGluThrLeuSerAspMet 454
161 ACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGAACAAATGGCCT 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
455 ThrGlyGlyLeuArgPheTyrAlaAsnLysAspLeu.....AsnSerLe 469
211 CATTGATGCTTTTGGGGCCCTTTCATCAGGAATGGAGCTGTCTCTCAGC 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
469 uileAspAlaPheSerArgIleSerSerThrSerGlySerValSerGlnG 486
261 GCTCCATCCAGCTTGAGAGTAAG 283
|||||:|||||:|||||:|||||:|||||:|||||:
486 InAlaLeuGlnLeuGluSerLys 493

seq_name: sp_rodent:Q9QX15

seq_documentation_block:
ID Q9QX15 PRELIMINARY; PRT; 902 AA.
AC Q9QX15;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CALCIUM-SENSITIVE CHLORIDE CONDUCTANCE PROTEIN-1.
GN CLCA1 OR MC1CA1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=WHOLE LUNG;
RA Gandhi R., Elble R.C., Gruber A.D., Schreuer K.D., Ji H.-L.,
RA Fuller C.M., Pauli B.U.;
RT "Molecular and functional characterization of a calcium-sensitive
chloride channel from mouse lung.";
RL J. Biol. Chem. 273:32096-32101(1998).
DR EMBL; AF047838; AAC79982.1; -.
DR MGD; MGI:1316732; Clcal
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 902 AA; 100107 MW; D9A915E698EA2D85 CRC64;

```

alignment_scores:
 Quality: 286.00 Length: 91
 Ratio: 3.620 Gaps: 1
 Percent Similarity: 86.813 Percent Identity: 65.934

alignment_block:
 US-09-049-696-7 x 088860 ..
 Align seg 1/1 to: 088860 from: 1 to: 901

US-09-049-696-7 x Q9QX15 ..
 Align seg 1/1 to: Q9QX15 from: 1 to: 902

```

11 ACTGATGATCTGAATTTGCTGCTGACGATGGGGAAGACACACTAT 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
405 ThrSerGlySerGluIleValLeuThrAspGlyGluAspAsnGlyI 421
161 AAGTGGGTGCTTTAAACGAGGTCAACAAAGTNGTGCCATCATCCACACAG 110
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
421 eArgSerCysPheGluAlaValSerArgSerGlyAlaIleIleHisThrI 438
111 TCAGTTTGGGGCCCTTCGACGCTCAAGAACTAGAGGAGCTGTCCAAATG 160
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
438 leAlaLeuGlyProSerAlaAlaArgGluLeuGluThrLeuSerAspMet 454
161 ACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGAACAAATGGCCT 210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
455 ThrGlyGlyLeuArgPheTyrAlaAsnLysAspLeu.....AsnSerLe 469
211 CATTGATGCTTTTGGGGCCCTTTCATCAGGAATGGAGCTGTCTCTCAGC 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
469 uileAspAlaPheSerArgIleSerSerThrSerGlySerValSerGlnG 486
261 GCTCCATCCAGCTTGAGAGTAAG 283
|||||:|||||:|||||:|||||:|||||:|||||:
486 InAlaLeuGlnLeuGluSerLys 493

seq_name: sp_mammal:O18742

seq_documentation_block:
ID O18742 PRELIMINARY; PRT; 794 AA.
AC O18742;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elble R.C., Widom J., Gruber A.D., Abdel-Chany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001262; AAB86530.1; -.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 794 AA; 88509 MW; B695E7256FC2C632 CRC64;

```

alignment_scores:
 Quality: 282.00 Length: 93
 Ratio: 3.439 Gaps: 1
 Percent Similarity: 88.172 Percent Identity: 60.215

alignment_block:
 US-09-049-696-7 x O18742 ..
 Align seg 1/1 to: O18742 from: 1 to: 794

```

11 ACTGATGATCTGAATTTGCTGCTGACGATGGGGAAGACACACTAT 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
406 ThrSerGlySerGluIleIleLeuThrAspGlyGluAspAsnGluI 422
61 AAGTGGGTGCTTTAAACGAGGTCAACAAAGTNGTGCCATCATCCACACAG 110
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
422 eAsnSerCysPheGluaspValLysArgSerGlyAlaIleIleHisThrI 439

```

```
111 TCGCTTTGGGCCCCCTGCAGCTCAAGAACTAGAGAGCTGTCTCAAAATG 160
.....|
439 leAlaLeuGlyProserAlaAlaLysGluLeuThrLeuSerAsnMet 455
161 ACAGGAGGTTTACAGACATGCTTCAGATCAAGTTCAAGAACATGGCCT 210
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
456 ThrGlyGlyTyrArgPhePheAlaAsnLysAspLe.....ThrGlyLe 470
211 CATTGATGCTTTGGGCCCCCTTCATCAGGAATGGAGCTGTCTCTCAGC 260
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
470 uThrAsnAlaPheSerArgIleSerArgSerGlySerIleThrGlnG 487

261 GCTCCATCCAGCTTCAGAGTAAGGGATTA 289
.....|
487 InAlaIleGlnLeuGluSerLysAlaLeu 496

seq_name: sp_mammal:018741
seq_documentation_block:
ID O18741 PRELIMINARY; PRT; 905 AA.
AC O18741;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001261; AAB86529.1; -
DR InterPro; IPR002035; VWFA.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00041; fn3; 1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR NCBI_TaxID=9913;
SQ SEQUENCE 905 AA; 101005 MW; D4CA6E7919115E88 CRC64;
[1]
```

seq_name: sp_mammal:018743

```
seq_documentation_block:
ID O18743 PRELIMINARY; PRT; 820 AA.
AC O18743;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001263; AAB86531.1; -
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR NCBI_TaxID=9913;
SQ SEQUENCE 820 AA; 91464 MW; 4E01330AEF5A1494 CRC64;
[1]
```

```
alignment_scores:
  Quality: 282.00 Length: 93
  Ratio: 3.439 Gaps: 1
Percent Similarity: 88.172 Percent Identity: 60.215

alignment_block:
US-09-049-696-7 x O18743 ..
```

Align seg 1/1 to: O18743 from: 1 to: 820

```
11 ACTGATGATCTGAAATGCTGCTGACGATGGGAGACAACTAT 60
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
406 ThrSerGlySerGluIleLeuLeuThrAspGlyGluAspAsnGluI 422
61 AAGTGGTGCTTTAACAGAGTCAACAAAGTNGTCCATCATCCACAG 110
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
422 eAsnSerCysPheGluAspValLysArgSerGlyAlaIleHisThrI 439
111 TCGCTTTGGGCCCCCTGCAGCTCAAGAACTAGAGAGCTGTCCAAATG 160
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
439 leAlaLeuGlyProserAlaAlaLysGluLeuThrLeuSerAsnMet 455
161 ACAGGAGGTTTACAGACATGCTTCAGATCAAGTTCAAGAACATGGCCT 210
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
456 ThrGlyGlyTyrArgPhePheAlaAsnLysAspLe.....ThrGlyLe 470
211 CATTGATGCTTTGGGCCCCCTTCATCAGGAATGGAGCTGTCTCTCAGC 260
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
470 uThrAsnAlaPheSerArgIleSerArgSerGlySerIleThrGlnG 487
```

```
261 GCTCCATCCAGCTTCAGAGTAAGGGATTA 289
.....|
487 InAlaIleGlnLeuGluSerLysAlaLeu 496
```

seq_name: sp_mammal:018741

```
seq_documentation_block:
ID O18741 PRELIMINARY; PRT; 905 AA.
AC O18741;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001261; AAB86529.1; -
DR InterPro; IPR002035; VWFA.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00041; fn3; 1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR NCBI_TaxID=9913;
SQ SEQUENCE 905 AA; 101005 MW; D4CA6E7919115E88 CRC64;
[1]
```

```
alignment_scores:
  Quality: 282.00 Length: 93
  Ratio: 3.439 Gaps: 1
Percent Similarity: 88.172 Percent Identity: 60.215

alignment_block:
US-09-049-696-7 x O18741 ..

Align seg 1/1 to: O18741 from: 1 to: 905
```

```
11 ACTGATGATCTGAAATGCTGCTGACGATGGGAGACAACTAT 60
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
406 ThrSerGlySerGluIleLeuLeuThrAspGlyGluAspAsnGluI 422
61 AAGTGGTGCTTTAACAGAGTCAACAAAGTNGTCCATCATCCACAG 110
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
422 eAsnSerCysPheGluAspValLysArgSerGlyAlaIleHisThrI 439
111 TCGCTTTGGGCCCCCTGCAGCTCAAGAACTAGAGAGCTGTCCAAATG 160
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
439 leAlaLeuGlyProserAlaAlaLysGluLeuThrLeuSerAsnMet 455
161 ACAGGAGGTTTACAGACATGCTTCAGATCAAGTTCAAGAACATGGCCT 210
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
456 ThrGlyGlyTyrArgPhePheAlaAsnLysAspLe.....ThrGlyLe 470
211 CATTGATGCTTTGGGCCCCCTTCATCAGGAATGGAGCTGTCTCTCAGC 260
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
470 uThrAsnAlaPheSerArgIleSerArgSerGlySerIleThrGlnG 487
261 GCTCCATCCAGCTTCAGAGTAAGGGATTA 289
.....|
487 InAlaIleGlnLeuGluSerLysAlaLeu 496
```

seq_name: sp_human:Q9UQC9

```
seq_documentation_block:
ID: Q9UQC9 PRELIMINARY; PRT; 943 AA.
AC Q9UQC9;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
```

```
alignment_scores:
  Quality: 218.00      Length: 90
  Ratio: 3.028         Gaps: 0
  Percent Similarity: 80.000  Percent Identity: 47.778
```

alignment block:

US-09-049-696-7 x 09Y6N2

Align seq 1/1 to: 09Y6N2 from: 1 to: 943

[illegible]

67 GTGCTTTTACGGAGGTCAACAAGAAGTGCCATCATCCACACAGTCGCTT
:
:
:
428 nCysLeuProThrValleuSerGlySerThrIleHisSerileAlaL 445

117 TGGGGCCCTCTGCGACGCTCAAGACTAGAGGAGCTGTCCAAATACGACGGA 166
||||| ||||||| ::||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
445 euGlySerSerAlaAalaProAsnLeuclucLuLeuSerArgLeuThrGly 461

440 euqysrseradadaprpasineuuuuuuueuseratglenurrgy
167 GGTTCAGACATATGCTTCCAGTAAGTTTCAGAACAATGGCCTCATTTA 216
||||| : : : : :
445 euqysrseradadaprpasineuuuuuuueuseratglenurrgy
462 GlyLeuIysPhePheValProAspLleSerAsnSerAsnSermMetIleas 478

462 GlyLeuIysPhePheValProAspIleSerAsnSerAsnSerMetIleAs 478
214 TGGTTTTGGGGCCCTTTCATCAGCAAAATGAGCTGTCTCTCACGGCTCCA 266

: 478 pAlaPheSerArgIleSerSerGlyThrGlyAspIlePheGlnGlnHisI 495
 . 267 TCCAGCTTGGAGTAGTAAAGGA 286
 . 495 leGlnLeuGluSerThrGly 501

495 leGlnLeuGluSerThrGly 501

seq_name: sp human:O9NXP1

seq documentation block.

```
seq_documentation_block:
ED  09NXP1  PRELIMINARY:  PRT:  469 AA
```

Q9NXP1;
AC
DT 01-OCT-2000 (TREFMPL) 15 (Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

CDNA FLJ20131 FIS, CLONE COL06357.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;

CX NCHL_TaxID=9606;

PP SEQUENCE FROM N.A.

RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Iku

RA Tañaka T., Nakamura Y., Isoqai T., Sugano S.;

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK000138; BAA90969.1; -
 SQ SEQUENCE 469 AA; 50694 MW; 6765986C886B3E56 CRC64;

alignment_scores:

Quality: 169.00 Length: 47
 Ratio: 3.930 Gaps: 0
 Percent Similarity: 91.489 Percent Identity: 70.213

alignment_block:

US-09-049-696-7 x Q9NXP1 ..

Align seg 1/1 to: Q9NXP1 from: 1 to: 469

149 CNGTCAAAATCACAGGAGTTTACAGACATATGCTCAGATCAAGTTCA 198

1 MetSerLysIleThrGlySerHisPheTyrValSerAspGluAlaG1 17

199 GAACAATGCCCTCATGTGCTTTGGGCCCTTTCATCAGGAATGGAG 248

17 nAsnAsnGlyLeuIleAspAlaPheGlyAlaLeuThrSerGlyAsnThra 34

249 CTGCTCTCAGCGCCATCCAGCTGAGAGTAAGGATTA 289

34 splSerGlnLysSerLeuGlnLeuGlySerGlyLeu 47

seq_name: sp_invertebrate:Q21317

seq_documentation_block:

ID Q21317 PRELIMINARY; PRT: 259 AA.

AC Q21317;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE COSMID K08D10.
 GN K08D10.9.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkison-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX Gelsel C., Bradshaw H.;
 RA Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX Waterston R.;
 RA Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U55857; AAA98027.1; -
 SQ SEQUENCE 259 AA; 28698 MW; 1C9CCAC9E93161E6 CRC64;

alignment_scores:

Quality: 79.50 Length: 102
 Ratio: 1.325 Gaps: 4

Quality: 80.00 Length: 100
 Ratio: 1.379 Gaps: 4
 Percent Similarity: 58.000 Percent Identity: 30.000

alignment_block:

US-09-049-696-7 x Q21317 ..

Align seg 1/1 to: Q21317 from: 1 to: 259

17 GCATCTGAAATGTGCTGCTGACG.....CATGGGAAGACAACTAT 60

52 GlySerValValIleLeuSerLysArgAsnProAsnGluAsnAsnI1 68

61 AAGTGGGTGCTTAAACGAGGTCAACAAAGTNGTGCCATCATCCACACAG 110

68 eSerAsnLeuValAlaLysIleArgSerHisHisGlyMetValHisVal1 85

111 TCGCT.....TTGGGGCCCTCTGCAGCTCAAGAACTAGAGGAG 148

85 leSerSerAsnThrProSerGlyGlySerGlnProLeuThrLeuTyrAsp 101

149 CTGTCCAAATCACAGGAGTTTACAGACATATGCTTCAGATCAAGTTCA 198

102 LeuSerPheLysThrAsnGlyLeuAlaAspPheArgAsnAspGlnPh 118

199 GAACAATGCCCTCATGTGCTTTGGG.....GCCCTT 233

118 eAsnAspSerGlnIleArgPhePheGlyLeuPheTrpProValAlaIle1 135

234 CATCAGGAATGGAGCTGCTCTCAGCGC...TCCATCCAGCTTGAGAGT 280

135 yrSerValAsnProValValSerGlyLysGlySerIleValLeuProser 151

seq_name: sp_invertebrate:Q18005

seq_documentation_block:

ID Q18005 PRELIMINARY; PRT: 505 AA.

AC Q18005;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOTHETICAL PROTEIN C15B12.4.
 GN C15B12.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX Nham M.;
 RT "The sequence of C. elegans cosmid C15B12.";
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX Waterston R.;
 RA Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U23529; AAK39164.1; -
 SQ SEQUENCE 505 AA; 51167 MW; 1F7BA5595AB69660 CRC64;

alignment_scores:

Quality: 79.50 Length: 102
 Ratio: 1.325 Gaps: 4

Percent Similarity: 58.824 Percent Identity: 26.471

alignment_block:

US-09-049-696-7 x Q18005

Align seg 1/1 to: Q18005 from: 1 to: 505

```
38  ACGATGGGGAAGACACACTATAAGTGGGTCTTAACGAGGTCAACA 87
   : : : : : : : : : : : : : : : : : : : : : : : :
208  SerAspGlnAsnIleThrIleSerGlyLeuValSerGlyLeuAsnSe 224
   : : : : : : : : : : : : : : : : : : : : : : : :
88  AAGTGTGCTCATCATCACACAGTCGCTTTGGGGCCCTCTCAGCTCAAG 137
   : : : : : : : : : : : : : : : : : : : : : : : :
224  rAsnGluGlyLeuValAsnThrLeuValLysGlyAsnGlyIleAsnG 241
   : : : : : : : : : : : : : : : : : : : : : : : :
138  AACTAGAGGAGCTGCCAAATGACAGA..... 166
   : : : : : : : : : : : : : : : : : : : : : : : :
241  lyThrAspGln.....LysMetThrGlyThrMetTyrGlyValAlaSer 255
   : : : : : : : : : : : : : : : : : : : : : : : :
167  .....GGTTACAGACATATGCTTCAGATCA 192
   : : : : : : : : : : : : : : : : : : : : : : : :
256  GlyLysGlyAsnSerThrLeuValGlyAlaSerSerIleValSerAsnGI 272
   : : : : : : : : : : : : : : : : : : : : : : : :
193  AGTTCAGAACAAATGGCTCATGATGCTTTTGGG.....GCCCTTT 233
   : : : : : : : : : : : : : : : : : : : : : : : :
272  nSerSerSerAlaglyGluIleGlnAlaPheGlyAsnSerAsnAlaPheS 289
   : : : : : : : : : : : : : : : : : : : : : : : :
234  CATCAGGAATGGAGCTGTC.....TCTCAGCGCTCCATCAGCTT 274
   : : : : : : : : : : : : : : : : : : : : : : : :
289  erAspGlyAsnSerSerValAsnLeuMetSerAsnThrAsnIleGluSer 305
   : : : : : : : : : : : : : : : : : : : : : : : :
275  GAGAGT 280
   : : : : : : : : : : : : : : : : : : : : : : : :
306  AspSer 307
```

seq_name: sp_bacteria:Q9ABQ3

seq_documentation_block:

```
ID  Q9ABQ3 PRELIMINARY; PRT; 687 AA.
AC  Q9ABQ3;
DT  01-JUN-2001 (TrEMBLrel. 17, Created)
DT  01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT  01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE  GENERAL SECRETION PATHWAY PROTEIN D.
GN  CC0173.
OS  Caulobacter crescentus.
OC  Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OX  NCBI_TaxID=69394;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=21173698; PubMed=11259647;
RA  Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA  Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA  Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA  DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA  Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty K.,
RA  Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA  Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT  "Complete genome sequence of Caulobacter crescentus.";
RL  Proc. Natl Acad. Sci. U.S.A. 98:4136-4141(2001).
DR  EMBL; A8005692; AKN22160.1; -
DR  TIGR; CC0173; -
KW  Complete proteome.
SQ  SEQUENCE 687 AA; 71703 MW; 1EEDBB7E4B11B563 CRC64;
```

alignment_scores:

Quality: 79.50 Length: 79
Ratio: 1.807 Gaps: 2
Percent Similarity: 55.696 Percent Identity: 31.646

alignment_block:

US-09-049-696-7 x Q9ABQ3

Align seg 1/1 to: Q9ABQ3 from: 1 to: 687

```
26  ATTGTGCTGCTGACGGATGGGAAGACAAACACT.....ATAAGTGGGTG 69
   : : : : : : : : : : : : : : : : : : : : : : : :
161  ValValValValAlaAspTyrAlaAspAsnValArgGlyIleArgGlyLe 177
   : : : : : : : : : : : : : : : : : : : : : : : :
70  CTTTAACGAGGTCAACAAAGTNGTCATCATCCACACAGTCGCTTTGG 119
   : : : : : : : : : : : : : : : : : : : : : : : :
177  uLeuAlaGlnValAspGlnAspArgAlaMetValHisThrValThrLeu 194
   : : : : : : : : : : : : : : : : : : : : : : : :
120  GGCCCTCTGCAGCTCAGAACTAGAGGAGCTGTCCAAATGACAGGAGGT 169
   : : : : : : : : : : : : : : : : : : : : : : : :
194  hrHisSerSerAlaArgGluIleAla..... 202
   : : : : : : : : : : : : : : : : : : : : : : : :
170  TTACAGACATATGCTTCAGATCAAGTTTCAGAACTAGCCCTCATTCATGC 219
   : : : : : : : : : : : : : : : : : : : : : : : :
203  .....GlnValLeuAsnAspMetLeuAlaThrPr 212
   : : : : : : : : : : : : : : : : : : : : : : : :
220  TTTTGGGGCCCTTTCATCATCAGAAATGGAGCTGTCTCT 256
   : : : : : : : : : : : : : : : : : : : : : : : :
212  oGlyAlaGluGlyLysSerGlyArgGlyAlaValThr 224
   : : : : : : : : : : : : : : : : : : : : : : : :
seq_name: sp_fungi:O14436
```

seq_documentation_block:

```
ID  O14436 PRELIMINARY; PRT; 728 AA.
AC  O14436;
DT  01-JAN-1998 (TrEMBLrel. 05, Created)
DT  01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT  01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE  CATALASE (EC 1.11.1.6).
GN  CAT1.
OS  Aspergillus fumigatus (Sartorya fumigata).
OC  Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC  Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX  NCBI_TaxID=5085;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CHUV 192-88;
RA  Calera J., Paris S., Monod M., Hamilton A.J., Debaupuis J.,
RA  Diaquin M., Lopez-Medrano R., Leal F., Latge J.;
RL  Infect. Immun. 65:0-0(0011).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CHUV 192-88;
RA  Latge J.;
RL  Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC  -!- FUNCTION: OCCURS IN ALMOST ALL AEROBICALLY RESPIRING ORGANISMS AND
CC  -!- SERVES TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN
CC  PEROXIDE.
CC  -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC  -!- COFACTOR: HEME GROUP.
DR  EMBL; U97574; AAB71223.1; -
DR  HSP; P21179; ICF9.
DR  InterPro; IPR002226; Catalase.
DR  Pfam; PF00199; catalase; 1.
DR  PRINTS; PR00067; CATALASE.
DR  ProDom; PD000510; Catalase; 1.
DR  PROSITE; PS00437; CATALASE_1; 1.
DR  PROSITE; PS00438; CATALASE_2; 1.
KW  Heme; Hydrogen peroxide; Iron; Oxidoreductase; Peroxidase.
FT  ACT_SITE 102 102 BY SIMILARITY.
SQ  SEQUENCE 728 AA; 79978 MW; 97F3EACCDFA5F47D CRC64;
```

alignment_scores:

Quality: 79.00 Length: 109
Ratio: 1.295 Gaps: 5
Percent Similarity: 55.963 Percent Identity: 31.193

alignment_block:

US-09-049-696-7 x 014436

Align seg 1/1 to: 014436 from: 1 to: 728

14 GATGATCTGAAATTTGGCTGG.....CTGACGGATGGGAAGA 51
 ||||| : : : : : |||||
 : : : : : |||||
 596 AspGlyValAspValValLeuValAlaGluArgLeuAlaAspGlyValAs 612
 : : : : : |||||
 : : : : : |||||
 52 CAACACATTAAGTGGGTGCTTTTAACGAGGTCAAAACAAGATNGTGCATCA 101
 : : : : : |||||
 : : : : : |||||
 612 pGlnThrTyrSer.....ThrSerAspAlaIleG 622
 : : : : : |||||
 : : : : : |||||
 1102 TCCACACAGTCGCTTTTGGGCCCTCTGCAGCTCAAGAACTAGAGAGCTG 151
 : : : : : |||||
 : : : : : |||||
 622 InPheAspAlaValValAlaAlaGlyAlaGluSerLeuPheAlaAla 638
 : : : : : |||||
 : : : : : |||||
 152 TCCAAATGACAGAGAGGTTTACAGACATATGCTTCA.....GATCAAGT 195
 : : : : : |||||
 : : : : : |||||
 639 SerSerPheThrGlyGlySerAlaAsnSerAlaSerGlyAlaSerSerIle 655
 : : : : : |||||
 : : : : : |||||
 196 TCAGAACAAATGCG.....CTCATGTAGTCTTT..... 223
 : : : : : |||||
 : : : : : |||||
 655 uTyrProThrGlyArgProLeuGlnIleLeuIleAspGlyPheArgPheG 672
 : : : : : |||||
 : : : : : |||||
 224GGGGCCCTTTTCATCAGGAATGGAGCTGTCTCTCAGCGCG 262
 : : : : : |||||
 : : : : : |||||
 672 lYlThrValGlyAlaLeuGlySerGlyThrAlaAlaLeuArgAsnAla 688
 : : : : : |||||
 : : : : : |||||
 263 TCCATCCAGCTTGAGTAAGGGATTA 289
 : : : : : |||||
 : : : : : |||||
 689 GlyIleAlaThrSerArgAspGlyVal 697

OM of: US-09-049-696-7 to: SwissProt_39:* out_format : pfs

Date: Mar 30, 2002 2:52 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+g2p.model -DEV=xlp  
-o=/cgn2_1/USPTO_spool/US09049696/runat_28032002_145238_2085/app_query.fasta_1.12579  
-DB=SwissProt_39 -OFMT=fastan -SUFFIX=rsp -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELCP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTEMT=pfs  
-NORM=ext -MINLEN=0 -MAXLEN=2000000000  
-USER=US09049696 -CGN1_1.165 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY  
-WAIT -THREADS=1
```

Search information block:

Query: US-09-049-696-7

Query length: 289

Database: SwissProt_39.*

Database sequences: 100059

Database length: 36664827

Search time (sec): 306.030000

score_list:

Sequence	Strd Orig	zScore	EScore	len	Documentation
SwissProt_39:ECIC_BOVIN	284.00	565.12	9.3e-25	903	P54281 bos taurus (bovine).
SwissProt_39:CATB_ASPFU	79.00	143.09	0.3692	728	Q92405 aspergillus fumigatus
SwissProt_39:TRPD_ACICA	74.00	138.98	1.30	349	P00500 acinetobacter calcoac
SwissProt_39:ATPD_ARATH	70.50	128.58	3.41	507	P92549 arabidopsis thaliana
SwissProt_39:ATPD_BRANA	70.50	128.58	3.41	507	P22201 brassica napus (rape)
SwissProt_39:ATPD_RAPSA	70.50	128.58	3.41	507	P23413 raphanus sativus (rad
SwissProt_39:PGK_MYCPN	68.50	126.27	5.69	409	P78018 mycoplasma pneumoniae
SwissProt_39:YAS9_SCHPO	66.50	117.13	10.17	738	Q10145 schizosaccharomyces
SwissProt_39:MYSS_CHICK	66.50	108.95	11.06	1938	P13538 gallus gallus (chick)
SwissProt_39:MYH8_HUMAN	66.00	107.92	12.62	1937	P13535 homo sapiens (human)
SwissProt_39:YABS_BACSU	65.50	124.41	12.05	245	P37561 bacillus subtilis. hy
SwissProt_39:SYD_AQAE	65.00	115.74	14.87	603	Q67589 aquifex acolicus. asp
SwissProt_39:GBLP_DROME	64.50	120.13	16.07	318	Q18640 drosophila melanogast
SwissProt_39:YPVV_MYCTF	64.50	117.17	16.56	451	P29578 methanobacterium ther
SwissProt_39:GLPK_MYCTU	64.50	116.01	16.76	517	P69664 mycobacterium tubercu
SwissProt_39:SYTC_SCHPO	64.00	112.37	19.65	703	P87144 schizosaccharomyces
SwissProt_39:RGP1_ASPAC	64.00	110.67	20.00	860	P48825 aspergillus aculeatus
SwissProt_39:MRP5_HUMAN	64.00	106.32	20.91	1437	Q15440 homo sapiens (human)
SwissProt_39:GSHI_PROMI	63.50	119.73	20.59	261	P31315 mus musculus (mouse).
SwissProt_39:HLVA_PROMI	63.50	104.49	24.06	1577	P16466 proteus mirabilis. h
SwissProt_39:HBB_SPECI	63.00	123.62	22.36	146	P09421 spermophilus citellus
SwissProt_39:MYH2_HUMAN	63.00	101.70	27.97	1941	Q9ukx2 homo sapiens (human)
SwissProt_39:MYSA_DROME	63.00	99.86	28.50	2411	P05661 drosophila melanogae
SwissProt_39:ATPD_BETVU	62.50	112.06	28.43	506	Q08735 beta vulgaris (sugar
SwissProt_39:ATPD_PHAVU	62.50	112.02	28.44	508	P03494 zea mays (maize). ath
SwissProt_39:ATPD_SOYBN	62.50	112.02	28.44	508	P24459 phaseolus vulgaris (k
SwissProt_39:ATPD_ORYSA	62.50	112.02	28.44	509	Q01915 glycine max (soybean)
SwissProt_39:ATPD_WHEAT	62.50	112.01	28.44	509	P15998 oryza sativa (rice).
SwissProt_39:ATPD_HELAN	62.50	111.99	28.45	510	P12862 triticum aestivum (wh
SwissProt_39:INUL_KLUMA	62.50	111.27	28.66	555	P18260 helianthus annuus (cd
SwissProt_39:PTRA_ECOLI	62.50	106.61	30.06	962	P28999 kluyveromyces marxian
SwissProt_39:RT31_ACTPL	62.50	105.88	30.28	1049	P55130 actinobacillus coli. pro
SwissProt_39:RT32_ACTPL	62.50	105.86	30.29	1052	P55130 actinobacillus pleur
SwissProt_39:Y619_TREPA	62.00	116.35	30.74	270	P83628 treponema pallidum. h
SwissProt_39:NQRA_CHLTR	62.00	111.74	32.22	465	P18282 chlamydia trachomatis
SwissProt_39:MYH1_HUMAN	62.00	99.64	36.47	1939	P12882 homo sapiens (human)
SwissProt_39:MYH3_CHICK	62.00	99.64	36.47	1940	P02565 gallus gallus (chick
SwissProt_39:HD_HUMAN	62.00	95.55	38.02	3144	P42858 homo sapiens (human).
SwissProt_39:ACF7_MOUSE	62.00	91.08	39.80	5327	Q9qgx0 mus musculus (mouse)

SwissProt_39:ILVH_BUCAI + 61.50 119.85 33.50 158 ! P57320 buchnera aphidicol
SwissProt_39:PLYD_PSEMA + 61.50 114.11 35.53 311 ! P27027 pseudomonas margin
SwissProt_39:CGGR_BACSU + 61.50 113.36 340 ! O32253 bacillus subtilis.
SwissProt_39:FLID_ECOLI + 61.50 110.67 36.80 467 ! P24216 escherichia coli.
SwissProt_39:VGLE_HSVBS + 61.50 108.91 37.47 575 ! Q08101 bovine herpesvirus

seq_name: SwissProt_39:ECIC_BOVIN

seq_documentation_block:

ID ECIC_BOVIN STANDARD; PRT; 903 AA.
AC P54281;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN (CALCIUM-ACTIVATED CHLORIDE
DE CHANNEL).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Trachea;
RX MEDLINE=96125078; PubMed=8537359;
RA Cunningham S.A., Awayda M.S., Buben J.K., Ismailov I.I.,
RA Arrate M.P., Berdiev B.K., Benos D.J., Fuller C.M.;
RT "Cloning of an epithelial chloride channel from bovine trachea.";
RL J. Biol. Chem. 270:31016-31026(1995).
CC !- FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE
CC SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME;
CC MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND
CC TRANSEPITHELIAL TRANSPORT.
CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC !- TISSUE SPECIFICITY: TRACHEA.
CC !- PTM: PHOSPHORYLATED BY CAM-KINASE II.
CC !- SIMILARITY: CONTAINS 1 VWFA DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U36445; AAC48511.1;
CC SMART; SM00327; VWA; 1.
CC PROSITE; PS50234; VWA; 1.
CC Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
CC Calcium channel; Phosphorylation; Glycoprotein.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 331 351 POTENTIAL.
FT TRANSMEM 617 637 POTENTIAL.
FT TRANSMEM 883 903 POTENTIAL.
FT DOMAIN 308 476 VWFA.
FT CARBOHYD 75 75 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 688 688 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 811 811 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 816 816 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 842 842 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 857 857 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 903 AA; 100305 MW; 9742AB1590908AEC CRC64;

alignment_scores:

Quality: 284.00 Length: 93

Ratio: 3.550 Gaps: 1
Percent Similarity: 86.022 Percent Identity: 62.366
alignment_block:
US-09-049-696-7 x ECLC_BOVIN ..
Align seg 1/1 to: ECLC_BOVIN from: 1 to: 903
11 ACTGATGGATCGAATTTGCTGCTGCGGATGGGGAAGCAACACTAT 60
|||||
405 ThrSerGlySerGluLeuLeuThrAspGlyGluAspAsnGluI 421
|||||
61 AAGTGGGTGCTTAAAGAGGTCAACAAAGTNGTCCATCATCCACAG 110
|||||
421 eHisSerCysIleGluGluValGlnSerGlyValIleHisThrI 438
|||||
111 TCCTTTGGGGCCCTCGAGCTCAAGACTAGAGGAGCTGCCAAATG 160
|||||
438 LealaLeuGlyProSerAlaAlaLysGluLeuGluThrLeuSerAspMet 454
|||||
161 ACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGAACAAATGGCCT 210
|||||
455 ThrGlyGlyHisArgPheTyrAlaAsnLysAspIle.....AsnGlyLe 469
|||||
211 CATGATGCTTTTGGGGCCCTTTCATCAGGAATGAGCTGTCTCTCAGC 260
|||||
469 uThrAsnAlaPheSerArgIleSerSerArgSerGlySerIleThrGlnG 486
|||||
261 GCTCCATCCAGCTTGAGAGTAAGGATTA 289
|||||
486 InThrIleGlnLeuGluSerLysAlaLeu 495
|||||

seq_name: SwissProt_39:CATB_AS PFU

seq_documentation_block:
ID CATB_AS PFU STANDARD; PRT; 728 AA.
AC Q92405; P79018;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CATALASE B (EC 1.11.1.6).
GN CATB
OS Aspergillus fumigatus (Sartoris fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5085;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 90240 / AF10;
RA Takasuka T., Anderson M., Denning D.W.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Wyssong D.R., Diamond R.D., Robbins P.W.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: OCCURS IN ALMOST ALL AEROBICALLY RESPIRING ORGANISMS AND
CC SERVES TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN
CC PEROXIDE.
CC -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC -!- COFACTOR: HEME GROUP.
CC -!- SIMILARITY: BELONGS TO THE CATALASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y07763; CAA69069.1; -;
DR EMBL; U87850; AAB48485.1; -;
DR HSSP; P21179; 11PH.

DR InterPro: IPR002226; Catalase.
DR Pfam: PF00199; catalase; 1.
DR PRINTS; PR00067; CATALASE.
DR PRODOM; PD000510; Catalase; 1.
DR PROSITE; PS00437; CATALASE_1; 1.
DR PROSITE; PS00438; CATALASE_2; 1.
KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
FT ACT_SITE 102 102 BY SIMILARITY.
FT ACT_SITE 175 175 BY SIMILARITY.
FT BINDING 389 389 PROXIMAL HEME LIGAND (BY SIMILARITY).
FT CONFLICT 35 35 A -> R (IN REF. 2).
FT CONFLICT 308 308 W -> R (IN REF. 2).
FT CONFLICT 595 595 DD -> EH (IN REF. 2).
FT CONFLICT 606 606 R -> A (IN REF. 2).
SQ SEQUENCE 728 AA; 79910 MW; 65E21979A63D85E7 CRC64;

alignment_scores:
Quality: 79.00 Length: 109
Ratio: 1.295 Gaps: 5
Percent Similarity: 55.963 Percent Identity: 31.193
alignment_block:
US-09-049-696-7 x CATB_AS PFU ..
Align seg 1/1 to: CATB_AS PFU from: 1 to: 728
14 CATGATCTGAAATTTGCTG.....CTGACGGATGGGAAGA 51
|||||
596 AspGlyValAspValValValAlaGluArgLeuAlaAspGlyValAs 612
52 CAACACTATAAGTGGTCTTTTAAAGAGGTCAACAAAGTNGTCCCATCA 101
|||||
612 pGlnThrTyrSer.....ThrSerAspAlaIleG 622
102 TCACACAGCTGCTTTGGGCCCTCTGCAGCTCAAGACTAGAGGAGCTG 151
|||||
622 InPheAspAlaValValAlaAlaGlyAlaGluSerLeuPheAlaAla 638
152 TCACAAATGACAGAGGTTTACAGACATATGCTTCA.....GATCAAGT 195
|||||
639 SerSerPheThrGlyGlySerAlaAsnSerAlaSerGlyAlaSerSerLe 655
196 TCAGAACAAATGCT.....CTCATGATGCTTTT..... 223
|||||
655 uTyrProThrGlyArgProLeuGlnIleLeuIleAspGlyPheArgPheG 672
224GGGGCCCTTTTCATCAGGAATGGAGCTGTCTCTCAGCGC 262
672 lYlYsThrValGlyAlaLeuGlySerGlyThrAlaAlaLeuArgAsnAla 688
263 TCATCCAGCTTCAGAGTAAGGATTA 289
689 GlyIleAlaThrSerArgAspGlyVal 697

seq_name: SwissProt_39:TRPD_ACICA

seq_documentation_block:
ID TRPD_ACICA STANDARD; PRT; 349 AA.
AC P00500;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.18).
GN TRPD.
OS Acinetobacter calcoaceticus.
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
OC Acinetobacter.
OX NCBI_TaxID=471;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88174326; PubMed=6599977;
RA Kaplan J.B., Goncharoff P., Seibold A.M., Nichols B.P.;

STRAIN=CV. COLUMBIA;
MDLXLINE=97141919; PubMed=8988169;
Unself M., Marlenfeld J.R., Brandt P., Brennicke A.;
"The mitochondrial genome of Arabidopsis thaliana contains 57 genes
in 366,924 nucleotides".
Nat. Genet. 15:57-61(1997).
[2]

RNA EDITING.
RA Marlenfeld J.R.;
Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
GRADIENT ACROSS THE MEMBRANE. THE ALPHA CHAIN IS A REGULATORY
SUBUNIT.
-!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
HAS THREE MAIN SUBUNITS: A, B AND C.
-!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
-!- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
-!- CAUTION: POSITIONS 431, 452 AND 495 ARE MODIFIED BY RNA EDITING.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch).

EMBL; Y08502; CAA69802.2; --
HSSP; P09219; ISKY.
DR SWISS-2DPAGE; P92549; ARATH.
DR InterPro; IPR000790; ATPase_A_C.
DR InterPro; IPR000194; ATPase_alpha_beta.
DR Pfam; PF004322; ATP-synt_A-C; 1.
DR Pfam; PF00006; ATP-synt_ab; 1.
DR ProDom; PD001099; ATPase_A_C; 1.
DR PROSITE; PS001152; ATPASE_ALPHA_BETA; 1.
KW ATP synthesis; CF(1); Hydrogen ion transport; Hydrolase; ATP-binding;
KW Mitochondrion; RNA editing.
DR NP_BIND 171 178 ATP (BY SIMILARITY).
FT ACT_SITE 373 373 BY SIMILARITY.
SQ SEQUENCE 507 AA; 55045 MW; 1927B3B8E1EFF51E CRC64;

alignment_scores:
Quality: 70.50 Length: 92
Ratio: 1.259 Gaps: 2
Percent Similarity: 60.870 Percent Identity: 25.000

alignment_block:
US-09-049-696-7 x ATP0_ARATH ..
Align seg 1/1 to: ATP0_ARATH from: 1 to: 507

23 GAATTTGCTGCTGACGGATGGGGAAGACACACTATAAGTGGGTCTT 72
|||||.....
53 GluMetValLeuPheAlaAsnGlyVallyGlyMetAlaLeuAsnLeuCl 69
|||||.....
78 TAACGAGGTCAAAACAAAGTNGGCCATCATCCACACACGCTTTGGGCG 122
|||||.....
69 uAsnGlu.....AsnValGlyIleValValPheGlyG 80
|||||.....
123 CCTCTGCAGCTCAACAACACTAGGAGCTGTGCCAAATGACAGAGGTTA 172
|||||.....
80 lyAspThrAlaIleIysGluGlyAspLeuVallyAspArgThrGlySerIle 96
|||||.....
173 CAGACATATGCTTTCAGATCAAGTTCAGAAACATGCCCTCATTTGATGCTTT 222
|||||.....
97 ValAspValProAlaGlyIlyAlaMetLeuGlyArgValValAspAlaMet 113
|||||.....
223 TGGGGCCCTTTTCATCAGGAATGAGGAGCTCTCTCT.....CAGCGCT 263
|||||.....


```

RN  SEQUENCE FROM N.A.
RP  STRAIN=972;
RA  Gentles S., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL  Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC  -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; Z69086; CAA93166.1;
DR  InterPro: IPR000504; RRM.
DR  Pfam; PF00076; RRM; 1.
DR  SMART; SM00360; RRM; 1.
DR  PROSITE; PS0102; RRM; 1.
DR  PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW  Hypothetical protein; RNA-binding.
FT  DOMAIN 363 434 RNA-BINDING (RRM).
SQ  SEQUENCE 738 AA; 82429 MW; 774B614B9D0A6023 CRC64;

alignment_scores:
  Quality: 66.50      Length: 98
  Ratio: 1.209       Gaps: 4
  Percent Similarity: 56.122      Percent Identity: 28.571

alignment_block:
US-09-049-696-7 x YAS9_SCHPO ..

Align seg 1/1 to: YAS9_SCHPO from: 1 to: 738

      8 CCAACTGATGATCTCAATTTGGCTGACGGATGGGAAGACAAAC... 55
      |||...: |||...: |||...: |||...: |||...: |||...:
      90 ProSerGlySerLeuGluProIleLeuAsnSerAlaThrSerAsn11 106
      56 .....ACTATAAGTGGGTGCTTTAACGAGGTCAACAAAGTNGTGCCA 98
      ||| ||| :|||...: |||...: |||...: |||...: |||...:
      106 eArgLeuThrProSerAspThrTyrAsnAsn117

      99 TCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGACTAGAGGAG 148
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
      118 .....ProValSerAspThrAsnAsnGluGlu 126

      149 CTGTCCAAAATGACAGAGGT.....TTACAGACATATGCTTCAGA 189
      :|||...: ||| :|||...: |||...: |||...: |||...: |||...:
      127 IleSerLysAsnIleTyrGlyAlaProIleLeuGluSerThrSerSerAs 143

      190 TCAAGTTCAGAACATGGCCTCATGATGCTTTTGGGGCCCTTCATCAG 239
      | :|||...: ||| :|||...: ||| :|||...: ||| :|||...: |||
      143 pHeGlnSerLysAspSerLeuSerThrThrGlnProSerValSerGlyG 160

      240 GAATGAGCTGCTCTCCAGCGCTCCATCCAGCTTGAGAGTA 281
      |||||...: |||||...: |||||...: |||||...: |||||...: |||||
      160 LysAsnGlySerThrSerGln.SerProSerLeuAspVal 173

seq_name: SwissProt_39: MYSS_CHICK

seq_documentation_block:
ID MYSS_CHICK STANDARD; PRT: 1938 AA.
AC F13538; O13228;
DT 01-JAN-1990 (Rel. 13, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN HEAVY CHAIN, SKELETAL MUSCLE, ADULT.
OS Gallus gallus (Chicken)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
```

```

OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEHORN; TISSUE=Pectoralis muscle;
RA Chao T.H., Bandman E., Moore L.;
RT "Cloning, nucleotide sequence and characterization of a full-length
RT cDNA encoding the myosin heavy chain from adult chicken pectoralis
RT major muscle.";
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-205.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041767; PubMed=1939027;
RA Hayashida M., Maizumi T., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: I.
RT Sequence of the amino-terminal 23 kDa fragment.";
RL J. Biochem. 110:54-59(1991).
RN [3]
RP SEQUENCE OF 206-636.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041768; PubMed=1939028;
RA Komine Y., Maizumi T., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: II.
RT Sequence of the 50 kDa fragment of subfragment-1.";
RL J. Biochem. 110:60-67(1991).
RN [4]
RP SEQUENCE OF 637-837.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041769; PubMed=1939029;
RA Maizumi T., Miyawaki T., Matsuzono K., Tanioka Y., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: III.
RT Sequence of the 22 kDa fragment and the alignment of the 23 kDa, 50
RT kDa, and 22 kDa fragments.";
RL J. Biochem. 110:68-74(1991).
RN [5]
RP SEQUENCE OF 838-1938.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041770; PubMed=1939030;
RA Maizumi T., Yajima E., Nagata S., Miyawaki T., Nakayama S., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: IV.
RT Sequence of the rod, and the complete 1,938-residue sequence of the
RT heavy chain.";
RL J. Biochem. 110:75-87(1991).
RN [6]
RP PRELIMINARY SEQUENCE OF 1-808.
RX MEDLINE=87092420; PubMed=3467365;
RA Maizumi T., Hayashida M., Tanioka Y., Komine Y., Matsuda G.;
RT "The primary structure of the myosin head.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:416-420(1987).
RN [7]
RP SEQUENCE OF 842-1270.
RX MEDLINE=90121764; PubMed=2610940;
RA Watanabe B.;
RT "Complete amino-acid sequence of subfragment-2 in adult chicken
RT skeletal muscle myosin.";
RL Biol. Chem. Hoppe-Seyler 370:1027-1034(1989).
RN [8]
RP SEQUENCE OF 852-1108.
RX MEDLINE=89374803; PubMed=2775482;
RA Watanabe B.;
RT "Amino-acid sequence of the short subfragment-2 in adult chicken
RT skeletal muscle myosin.";
RL Biol. Chem. Hoppe-Seyler 370:549-558(1989).
RN [9]
RP SEQUENCE OF 1145-1270.
RX MEDLINE=89228549; PubMed=2713098;
RA Watanabe B.;
RT "Amino-acid sequence of the hinge region in chicken myosin
RT subfragment-2.";
RL Biol. Chem. Hoppe-Seyler 370:55-61(1989).
RN [10]
RP SEQUENCE OF 1857-1938 FROM N.A.
RX MEDLINE=87217964; PubMed=3034534;
```

```

ET MOD_RES 755 755 METHYLATION (MONO-).
ET MOD_RES 697 697 ALKYLATION (SH-1).
ET MOD_RES 707 707 ALKYLATION (SH-2).
ET CONFLICT 907 907 C -> Q (IN REF. 7 AND 8).
ET CONFLICT 980 980 L -> F (IN REF. 1).
ET CONFLICT 1343 1343 E -> D (IN REF. 5).
ET CONFLICT 1545 1545 S -> A (IN REF. 5).
ET CONFLICT 1796 1797 HV -> QL (IN REF. 5).
ET CONFLICT 1830 1830 S -> A (IN REF. 5).
ET CONFLICT 1863 1863 I -> V (IN REF. 10).
ET CONFLICT 1929 1931 IHG -> FH (IN REF. 10).
SQ SEQUENCE 1938 AA; 223013 MW; EDDOICEA2681E10F CRC64;

alignment_scores:
  Quality: 66.50      Length: 79
  Ratio: 1.330        Caps: 4
Percent Similarity: 63.291 Percent Identity: 31.646

alignment_block:
US-09-049-696-7 x MYSS_CHICK ..
Align seg 1/1 to: MYSS_CHICK from: 1 to: 1938

11 ACTGATGATCTGAATATGTGCTGCTGACGGATGGGGAAGACAACACTACT 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
159 SerAspAsnAlaTyrGlnPheMetLeuThrAspArgGluAsnGlnSerI 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 A.....ATGGGTGCTTTAACGACGAGTCAACAAAGACTNGTG 95
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
175 eLeuIleThrGlyGluSerGlyAlaGlyLysThrValAsnThrLysArgV 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
96 CCATCATCCACACAGTCGGCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAG 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
192 allIeGlnTyrPheAlaThrIleAlaAlaSerGlyGluLysLysGlu 208
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
146 GAGCTGTCC...AAAATGACAGGAGGTTTACAGACATATGCTTCAGATCA 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
209 GluGlnSerGlyLysMetGlnGlyThrLeuGlu.....AspG1 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
195 A...GTTTCAGAACAANTGGCTCATTCATGCTTTTGGG 226
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
221 nIleIleSerAlaAsnProLeuLeuGluAlaPheGly 233

seq_name: SwissProt_39:MYH8_HUMAN

seq_documentation_block:
ID MYH8_HUMAN STANDARD; PRT; 1937 AA.
AC F13535; Q14910;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN HEAVY-CHAIN, SKELETAL MUSCLE, PERINATAL (MYHC-PERINATAL).
GN MYH8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
QC NCBI_TaxID=9606;
RL 144
SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=95324556; PubMed=2373371;
RA Karsch-Mizrahi I., Feghali R., Shows T.B. Jr., Leinwand L.A.;
RT "Generation of a full-length human perinatal myosin heavy-chain-
RT encoding cDNA.";
RL Gene 89:289-294(1990).
[2]
SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=95324556; PubMed=7601129;
RA Juilian E.H., Kelly A.M., Pompidou A.J., Hoffman R., Schiaffino S
RA Stédman H.H., Rubinstein N.A.;
RT "Characterization of a human perinatal myosin heavy-chain

```

transcript.";

Eur. J. Biochem. 230:1001-1006(1995).

[3]

SEQUENCE OF 502-1937 FROM N.A.

TISSUE=Skeletal muscle;

MEDLINE=90235862; PubMed=1691980;

Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,

Arnold H.H.;

"Identification of three developmentally controlled isoforms of human

myosin heavy chains";

Eur. J. Biochem. 189:55-65(1990).

[4]

SEQUENCE OF 860-1937 FROM N.A.

MEDLINE=89234168; PubMed=2715179;

Feghali R., Leinwand L.A.;

"Molecular genetic characterization of a developmentally regulated

human perinatal myosin heavy chain.";

J. Cell Biol. 108:1791-1797(1989).

[5]

SEQUENCE OF 1-46 FROM N.A.

Esser K., Tidhar A., Myszkowski M.;

"Isolation and characterization of the human perinatal MHC promoter.";

Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.

-!- FUNCTION: MUSCLE CONTRACTION.

-!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2

HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)

AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).

-!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.

-!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING

CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,

CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

-!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY

ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.

-!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT

MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE

SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED

SUBFRAGMENT (S2).

-!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

-!- SIMILARITY: CONTAINS 1 IO DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch).

EMBL; M36769; AAC17185.1; -;

EMBL; Z38133; CAA86293.1; -;

EMBL; X51592; CAA35941.1; -;

EMBL; M32520; AAA36346.1; -;

EMBL; AF067143; AAC21557.1; -;

PIR; A30220; A30220.

HSP; P13538; 2MYS.

MIM; 160741; -;

InterPro; IPR000048; IQ.

InterPro; IPR002928; Myosin_tail.

InterPro; IPR001609; myosin_head.

Pfam; PF00612; IQ; 1.

Pfam; PF00063; myosin_head; 1.

Pfam; PF01576; Myosin_tail; 1.

PRINTS; PR00193; MYOSINHEAVY.

ProDom; PD000355; myosin_head; 1.

SMART; SM00015; IQ; 1.

SMART; SM00242; MYSC; 1.

PROSITE; PS50096; IQ; 1.

Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;

ATP-binding; Methylation; Alkylation; Multigene family;

Calmodulin-binding.

DOMAIN 1 780 MYOSIN HEAD-LIKE.

FT 781 813 IQ.

DOMAIN 842 1937

FT COILED COIL (POTENTIAL).

FT

CC *Protophila melanogaster* (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

DR PROSITE; PS50082; WD_REPEATS_2; 6.
ID PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT REPEAT 13 44 WD 1.
FT REPEAT 62 92 WD 2.
FT REPEAT 104 134 WD 3.
FT REPEAT 147 179 WD 4.
FT REPEAT 191 221 WD 5.
FT REPEAT 232 261 WD 6.
FT REPEAT 282 312 WD 7.
FT CONFLICT 104 104 G -> R (IN REF. 1).
SQ SEQUENCE 318 AA; 35618 MW; 202930BE51AE822 CRC64;

alignment_scores:
    Quality: 64.50      Length: 105
    Ratio: 1.402        Gaps: 6
    Percent Similarity: 43.810      Percent Identity: 24.762

alignment_block:
US-09-049-896-7 x GBLP_DROME ..
Align seg 1/1 to: GBLP_DROME from: 1 to: 318

      7   TCCAAATGATGGATTCGAATTTGCTGCTGACGGATGGGGAAGACAACA 56
      ||||| ||| ||| ||||| |||
162  SerAsnProIleIleValSerCys.....GlyTirAspArgThrVa 175

      57  CTATAAGTGGTGCTTTACGAGGTCACAAAGTNGTCCCATCATCCAC 106
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175  lLysValTrpAsnLeuAlaAsnCysLysLeuLysAsnAsnHisHisGlyH 192

      107 ACAGTCGCTTT.....GGGCGCCTCTGC 129
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
192  isAsnGlyTyrLeuAsnThrValThrValSerProAspGlySerLeuCys 208

      130 AGCTCAAGAATACTACGAGAGCT..... 150
      ||||| ||| ||| |||
209  ThrSerGlyGlyLysAspSerLysAlaLeuLeuTrpAspLeuAsnAspGl 225

      151 .....GTCCAAATGACAGGAGGTTTACAGACATATGCT 184
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
225  yLysAsnLeuTyrThrLeuGluHisAsnAspIleIleAsnAlaLeuCysP 242

      185 TCAGATCAAGTTCAGACAATGCGCTCATGTAGTGCTT...TGGGGCCCT 231
      || ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
242  he...SerProAsnArgTyrTrpLeu.....CysValAlaTyrGlyPro 255

      232 TTCATCAGGAATGG 246
      ||||| |||
256  SerIleLysIleTrp 260

seq_name: SwissProt_39:YPVB_METTF

seq_documentation_block:
ID YPVB_METTF STANDARD; PRT; 451 AA.
AC P29578;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE HYPOPHETICAL 49.6 KDA PROTEIN (ORF5B).
OS Methanobacterium thermoformicum.
CS Plasmid pFV1.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteri
CC Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 3848 / THF;
RX MEELINE-93126090; PubMed=1336177;
RA Noelling J., van Beden F.J.M., Eggen R.I.L., de Vos W.M.;
RT "Modular organization of related Archaeal plasmids encoding
PT restriction-modification systems in Methanobacterium

```



```

RT thermoformicum."
RL Nucleic Acids Res. 20:6501-6507(1992).
CC -1- SIMILARITY: TO ORF5 IN PF21.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X68366; CAA48430.1; -
DR PIR: S26441; S26441.
DR InterPro: IPR001434; DUF11.
DR Pfam: PF01345; DUF11; 2.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 451 AA; 49670 MW; 9949C4FF79119BAF CRC64;

alignment_scores:
  Quality: 64.50 Length: 72
  Ratio: 1.290 Gaps: 1
  Percent Similarity: 69.444 Percent Identity: 25.000

alignment_block:
US-09-049-696-7 x YPVB_METTF ..
Align seg 1/1 to: YPVB_METTF from: 1 to: 451
62 AGTGGGTCCTTTAAGAGGTCACAAAGTNGTCCCATCTCCACAGT 111
149 AlGlyTyrPheAspAlaThrArgLutrp...IlePheGluThrLe 164
112 CGCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAATGA 161
164 uAlaAlaGlyGluGluAlaThrLeuThrLeuAspCysLeuValAsnArgT 181
162 CAGAGGTTTACAGACATAGCTTCAGATCAAGTTCAGAACATGGCCTC 211
181 hrGlyGluLeuSerAspTyrValSerValArgGluValAspGluGlyAsp 197
212 ATTGATGCTTTTGGGGCCCTTCATCAGGAATGGAGCTGTCTCTCAGCG 261
198 ValAsnValTyrAsnAsnMetAlaHisAlaSerValAlaValLysGlyTh 214
262 CTCCATCCAGCTTGAG 277
214 rAspLeuAspLeuAsp 219

seq_name: SwissProt_39:GLPK_MYCTU

seq_documentation_block:
ID GLPK_MYCTU STANDARD; PRT; 517 AA.
AC O69664;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL 3-PHOSPHOTRANSFERASE)
DE (GLYCEROKINASE) (GK).
GN GLPK OR RV3696C OR WT3798 OR MTW025.044C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

```

```

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver K., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton J., Squares R., Squares S.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of mycobacterium tuberculosis from the
RT complete genome sequence.";
RN Nature 393:537-544(1998).
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Deicher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: KEY ENZYME IN THE REGULATION OF GLYCEROL UPTAKE AND
CC METABOLISM.
CC -1- CATALYTIC ACTIVITY: ATP + GLYCEROL = ADP + GLYCEROL 3-PHOSPHATE.
CC -1- PATHWAY: RATE-LIMITING STEP IN GLYCEROL UTILIZATION.
CC -1- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUCONOKINASE /
CC GLYCEROKINASE / XYLULOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL022121; CAA18018.1; -
DR EMBL: AE007177; AAK48165.1; -
DR HSP: P08859; 1GLB.
DR TIGR: MT3798; -
DR Tuberculist; RV3696C; -
DR InterPro: IPR000577; FGGY_kin.
DR Pfam: PF00370; FGGY; 1.
DR PROSITE: PS00445; FGGY_KINASES.2; 1.
DR PROSITE: PS00933; FGGY_KINASES.1; 1.
KW Glucitol metabolism; Transferase; Kinase; ATP-binding;
KW Complete proteome.
FT NPI_BIND 165 177 ATP (PROBABLE).
SQ SEQUENCE 517 AA; 55859 MW; 7E3F30DCF2A1E63E CRC64;

alignment_scores:
  Quality: 64.50 Length: 90
  Ratio: 1.132 Gaps: 4
  Percent Similarity: 63.333 Percent Identity: 30.000

alignment_block:
US-09-049-696-7 x GLPK_MYCTU ..
Align seg 1/1 to: GLPK_MYCTU from: 1 to: 517
17 GGATCTCAATTTGGCTGCTGACGGATGGGAGACACACACTATAAGTGG 66
292 GlyCluThrIleValArgSerAsnGlyLeuLeuThrThrValCysTy 308
67 GTGCTTTAAGAGTCAACAAAGTNGTGCATCATCCACACAGTCGCTT 116
308 rGlnPheGlyAsnAlaValGlnTrpLeuArgAspGlnLeuGlyIleIleSer 341
117 TGGGGCCCTCTGCAGCTCAACAACATA...GAGGAGCTGTCCAAATGACA 163
325 alThrGlySerAlaValGlnTrpLeuArgAspGlnLeuGlyIleIleSer 341
164 GGAGGTTTACAGACATATGCT...TCAGTCAAGTTCAGACACATGGC... 208
342 GlyAlaAlaGlnSerGluAlaLeuAlaArgGlnValProAspAsnGlyG 358

```

209CTCAATTCATGCTTTTGGGCCCTT.....T 233
 358 yMetTyrPheValProAlaPheSerGlyLeuPheAlaProTyrTyrArgS 375
 234 CATCAGGAATGGAGCTGC 253
 375 erAspAlaArgGlyValaile 381

OM of: US-09-049-696-7 to: PIR_68.* out_format : pfs

Date: Mar 30, 2002 2:26 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlp
-Q/cgn2_1/uspto_spool/US09049696/runat_28032002_145238_2031/app_query.fasta_1.12579
-DB=PIR_68 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THRM_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USBR=US09049696_@CGN1_1_367 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-049-696-7
Query length: 289
Database: PIR_68.*
Database sequences: 219241
Search length: 7617452
Search time (sec): 470.790000

score.list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
PIR2:T02205	+	282.00	7.9e-25	905	Lu-ECAM-1 protein - bovine
PIR2:T29382	+	80.00	0.4192	259	hypothetical protein K08D10.9 -
PIR2:T15502	+	79.50	0.5105	494	hypothetical protein C15B12.4 -
PIR2:T15071	+	77.50	0.8465	337	hypothetical protein R07C12.1 -
PIR2:T22258	+	77.50	0.8797	505	hypothetical protein F46A8.2 -
PIR2:T22135	+	76.50	1.43.74	450	probable fmu protein - Pyrococ
PIR2:T22264	+	74.50	1.43.04	298	hypothetical protein F46A8.1 -
PIR2:NPKEDE	+	74.00	140.63	349	anthranilate phosphoribosyltran
PIR2:T15069	+	72.50	137.24	358	hypothetical protein R07C12.3 -
PIR2:F84811	+	71.00	128.48	528	probable retroelement pol poly
PIR1:PWRA1	+	70.50	130.05	587	H+-transporting ATP synthase (F
PIR2:S12309	+	70.50	130.05	507	H+-transporting ATP synthase (E
PIR2:T15070	+	70.50	129.90	516	hypothetical protein R07C12.2 -
PIR2:F96584	+	69.50	133.84	7.22	hypothetical protein F20D21.13
PIR2:T37223	+	69.50	129.52	7.56	probable secreted protein - Str
PIR2:S54167	+	68.50	124.74	13.08	set1A protein - Shigella flexne
PIR2:T37338	+	68.50	127.65	9.89	phosphoglycerate kinase (EC 2.7
PIR2:T29018	-	68.50	121.91	10.54	hypothetical protein ZK84.1 - C
PIR2:C36530	+	68.00	127.98	11.15	hypothetical protein F13F21.18
PIR2:T39040	+	68.00	117.73	12.50	hypothetical protein SPAC6F6.0
PIR2:D75033	+	67.50	124.74	13.08	probable proliferating-cell nuc
PIR2:T28698	+	67.50	110.74	15.29	hypothetical protein - Streptoc
PIR2:G72512	-	67.00	134.54	13.27	hypothetical protein APE2076 -
PIR2:T20145	+	67.00	110.23	17.40	hypothetical protein C52A11.4
PIR1:JC5394	+	66.50	130.02	15.79	DJ-1 protein - human
PIR2:H72701	+	66.50	122.57	17.16	hypothetical protein APE1031 -
PIR2:T38767	+	66.50	118.39	17.98	probable RNA-binding protein -
PIR2:S36111	+	66.50	117.93	18.07	osteoblast-specific factor 2 -
PIR2:S36110	+	66.50	117.32	18.20	osteoblast-specific factor 2 -
PIR1:JX0178	+	66.50	110.14	19.71	myosin heavy chain, fast skele
PIR2:S49119	+	66.00	119.52	20.09	embryonic/neonatal myosin heavy
PIR2:T38055	+	66.00	109.09	22.57	myosin heavy chain, perinatal
PIR2:S66095	+	65.50	125.69	21.23	conserved hypothetical protein
PIR2:T09632	+	65.50	121.90	22.14	ygap protein - Lactobacillus de
PIR2:D86807	+	65.50	117.54	23.25	hypothetical protein pthA (impc
PIR2:G75121	+	65.00	122.60	24.87	nucleolar protein PAB0523 - Pyr
PIR2:H71009	+	65.00	122.49	24.90	probable proliferating-cell nuc
PIR2:D70445	+	65.00	116.94	26.49	aspartate--tRNA ligase (EC 6.1.
PIR2:G82176	+	64.50	124.81	27.45	probable outer membrane protein
PIR2:T40193	+	64.50	119.29	29.20	hypothetical protein SPBC30D10.
PIR2:S26441	+	64.50	118.37	29.50	hypothetical protein 5b - Metha

PIR2:A70793 + 64.50 117.20 29.89 517 ! probable glycerol kinase - M
PIR2:S77625 + 64.50 111.59 31.82 997 ! mannuronan C-5-epimerase (EC
PIR2:A56559 + 64.50 110.48 32.22 1136 ! enhancer-trap-locus-1 prote

seq_name: pir2:T02205

seq_documentation_block:

Lu-ECAM-1 protein - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 01-Dec-2000
C;Accession: T02205; T02152; T02171
R;Eblle: R.C.; Widom, J.; Gruber, A. D.; Abdel-Ghany, M.; Levine, R.; Goodwin, A.; Pau
Submitted to the EMBL Data Library, April 1997
A;Description: Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial
A;Reference number: Z14590
A;Accession: T02205
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-905 <ELB>
A;Cross-references: EMBL:AF001261; NID:g2623762; PIDN:AAB86529.1; PID:g2623763
A;Experimental source: lung
A;Accession: T02152
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-793, 'SGSFP', 'M', '862', 'RF', '865-866', 'Q', '868', 'AKVLEIQ', '876', 'QHQ', '880', 'FQ', '880', 'FQ'
A;Cross-references: EMBL:AF001263; NID:g2623766; PIDN:AAB86531.1; PID:g2623767
A;Experimental source: lung
A;Accession: T02171
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-792, 'ES', '<EL3>
A;Cross-references: EMBL:AF001262; NID:g2623764; PIDN:AAB86530.1; PID:g2623765
A;Experimental source: lung

alignment_scores:
Quality: 282.00 Length: 93
Ratio: 3.439 Gaps: 1
Percent Similarity: 88.172 Percent Identity: 60.215

alignment_block:

US-09-049-696-7 x T02205

Align seg 1/1 to: T02205 from: 1 to: 905

11 ACTGATGGATCTGAATTTGCTGCTGACGATGGGAGACACAACTAT 60
|||||
406 ThrSerGlySerGluIleLeuLeuThrAspGlyGluAspAsnGluI 422
61 AAGTGGTCTTTAAGCAGGCTCAACAACTGTCCTCATCCACACAG 110
|||||
422 easnSerCysPheGluAspVallylArgSerGlyAlaIleHisThrI 439
111 TCGCTTTGGGGCCCTCTGAGCTAGAACTAGAGAGCTGCCAAATG 160
439 lcaIaLeuGlyProserAlaAlaLysGluLeuGluThrLeuSerAsnMet 455
161 ACAGAGAGTTTACAGACATATGCTTCAGATCAAGTTCAGAACTGGCCT 210
456 ThrGlyGlyTyArgPhePheAlaAsnLysAspIle.....ThrGlyLe 470
211 CATTGATGCTTTGGGGCCCTTTTCATCAGAAATGAGCTGCTCTCAGC 260
470 uThrAsnAlaPheSerArgIleSerSerArgSerGlySerIleThrGlnG 487
261 GCTCCATCCAGCTTCAGACTAAGGGATTA 289
487 InAlaIleGlnLeuGluSerLysAlaLeu 496

seq_name: pir2:T29382

seq_documentation_block:

hypothetical protein K08D10.9 - Caenorhabditis elegans

Tue Apr 2 09:40:12 2002

C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Jun-2000
 C:Accession: T29382
 R:Geisler, C.; Bradshaw, H.
 submitted to the EMBL Data Library, April 1996
 A:Description: The sequence of C. elegans cosmid K08D10.
 A:Reference number: Z20616
 A:Accession: T29382
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-259 <GEN>
 A:Cross-references: EMBL:U05957; PIDN:AAA98027.1; GSPDB:GNO0022; CESP:K08D10.9
 A:Experimental source: strain Bristol N2; clone K08D10
 C:Genetics:
 A:Gene: CESP:K08D10.9
 A:Map position: 4
 A:Introns: 40/3; 120/3; 175/1; 208/3; 249/2
 C:Superfamily: Caenorhabditis hypothetical protein R07C12.3

alignment_scores:
 Quality: 80.00 Length: 100
 Ratio: 1.379 Gaps: 4
 Percent Similarity: 58.000 Percent Identity: 30.000

alignment_block:
 US-09-049-696-7 x T29382 ..

Align seg 1/1 to: T29382 from: 1 to: 259

```

17 GGATCTGAATGCTGCTGACG.....GATGGGAAGACACACTAT 60
|||||
52 GlySerValValIleValIleSerLysArgAsnProAsnGluAsnIle 68
|||||
61 AAGTGGGTCCTTAACAGAGTCAACAAAGTGTGCCATCATCCACACAG 110
|||||
68 eSerAsnLeuValAlaLysIleArgSerHisHisGlyMetValHisValI 85
|||||
111 TCGCT.....TTGGGGCCCTCTGCAGCTCAAGAACTAGAGGAG 148
|||||
85 leSerSerAsnThrProSerGlyGlySerGlnProLeuThrLeuTyrrAsp 101
|||||
149 CTGTCCAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTCA 198
|||||
102 LeuSerPheLysThrAsnGlyLeuAlaAspPheArgAsnAspGlnPh 118
|||||
199 GAACATGCCCTCATGTGCTTTGGG.....GCCCTTT 233
|||||
118 eAsnAspSerGlnIleArgPheGlyLeuPheTrpProValAlaIleT 135
|||||
234 CATCAGGAATGGAGCTGTCTCTCAGCGC...TCCATCAGCTTCAGAGT 280
|||||
135 yrSerValAsnProValSerGlyLysGlySerIleValLeuProSer 151
|||||

```

seq_name: pir2:T15502

seq_documentation_block:
 hypothetical protein C15B12.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T15502
 R:Nhan, M.
 submitted to the EMBL Data Library, March 1995
 A:Description: The sequence of C. elegans cosmid C15B12.
 A:Reference number: Z18362
 A:Accession: T15502
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-494 <NHAS>
 A:Cross-references: EMBL:U23529; NID:g746592; PID:g746596; PIDN:AAC46579.1; CESP:C15B12.
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:C15B12.4

A:Introns: 68/3; 104/1; 140/3; 167/1; 228/1; 294/1; 322/3; 357/1; 386/1; 422/3; 449/1

alignment_scores:
 Quality: 79.50 Length: 102
 Ratio: 1.325 Gaps: 4
 Percent Similarity: 58.824 Percent Identity: 26.471

alignment_block:
 US-09-049-696-7 x T15502 ..

Align seg 1/1 to: T15502 from: 1 to: 494

```

38 ACGGATGGGGAAGACACACACTATTAAGTGGTGTTCACGAGGTCAACA 87
|||||
224 SerAspAspGlnAsnIleThrIleSerGlyLeuValSerGlyLeuAsnSe 240
|||||
88 AAGTNGTCCCATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAG 137
|||||
240 rAsnGluGlyLeuValAsnThrLeuValLysGlyAsnGlyIleIleAsnG 257
|||||
138 AACTAGAGGAGCTGTCCAAATGACAGCA..... 166
|||||
257 lyThrAspGln.....LysMetThrGlyThrMetTyrGlyValAlaSer 271
|||||
167 .....GGTTACAGACATATGCTTCAGATCA 192
|||||
272 GlyLysGlyAsnSerThrLeuValGlyAlaSerSerIleValSerAsnGI 288
|||||
193 AGTTCAGACAAATGGCTCATGATGCTTTGGG.....GCCCTTT 233
|||||
288 nSerSerAlaGlyGluIleGlnAlaPheGlyAsnSerAsnAlaPheS 305
|||||
234 CATCAGGAATGGAGCTGTCTCTCAGCGC...TTCAGCGCTCCATCCAGCTT 274
|||||
305 eAspGlyAsnSerSerValAsnLeuMetSerAsnThrAsnIleGluSer 321
|||||
275 GAGAGT 280
|||||
322 AspSer 323

```

seq_name: pir2:T15071

seq_documentation_block:
 hypothetical protein R07C12.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jan-2000
 C:Accession: T15071
 R:Geisler, C.; Beck, C.; O'Brien, D.
 submitted to the EMBL Data Library, December 1997
 A:Description: The sequence of C. elegans cosmid R07C12.
 A:Reference number: Z18286
 A:Accession: T15071
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-337 <GEN>
 A:Cross-references: EMBL:AF038622; NID:g2736331; PID:g2736332; PIDN:AAB94150.1; GSPDB
 A:Experimental source: strain Bristol N2; clone R07C12
 C:Genetics:
 A:Gene: CESP:R07C12.1
 A:Map position: 4
 A:Introns: 92/2; 123/3; 203/3; 255/1; 286/3; 327/2
 C:Superfamily: Caenorhabditis hypothetical protein R07C12.3

alignment_scores:
 Quality: 77.50 Length: 95
 Ratio: 1.384 Gaps: 4
 Percent Similarity: 58.947 Percent Identity: 26.316

alignment_block:
 US-09-049-696-7 x T15071 ..

Align seq 1/1 to: T15071 from: 1 to: 337

[illegible]

seq name: pir2:T22258

```
seq documentation block:
```

hypothetical protein F46A8.2 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22258
R:Harris, B.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19539
A:Accession: T22258
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1505 <WIL>
A:Cross-references: EMBL:Z81539; PIDN:CA804388.1; GSPDB:GN00019; CESP:F46A8.2
A:Experimental source: clone F46A8

```

alignment_scores:
  Quality: 77.50
  Ratio: 1.409
  Percent Similarity: 55.000
  Length: 100
  Gaps: 3
  Percent Identity: 26.000

alignment_block:
  US-09-049-696-7 x T32258

```

Align seq 1/1 to: T22258 from: 1 to: 505

2 AAATATCCAACCTGATGGATCGAAATTGTGCTGCTGACG.....GATGG 45
 343 GlnPheSerThrGlySerGlyLeuValLeuLeuLeuSerLysArgAsnPr 359
 46 GGAAGACAACACATATAGTGGTCTTTAAACGAGGCTCAACAAAGTNGTG 95
 359 oAspGluAsnAspLysLeuLeuValArgSerHisHisG 376
 96 CCATCATCCACAGTCGCT.....TTGGGGCCCTTCGACGT 133
 376 lyMetValHisValIleAlaSerAsnThrProSerGlyClySerGlnPro 392
 134 CAAAGACTAGAGGAGCTGTCAAATGACAGGAGGTTTACACACATATGCG 183
 393 LeuThrLeuTyrAspLeuSerSerLysThrAsnGlyLeuAlaAspPheAr 409

```

184 TTCAGATCAAGTTGACAACAATGGCCTCATTTGAT..... 217
      |||:::      :::::|||||
409 gileAspSerGluPheAspGluSerValIleProPheAlaGlyLeuAsnT 426
      .....GCTTTGGGGCCCTTCATCAGGNAATGGAGCT 250
      ||| ::| ||| ||||| |||||:::
426 yrProIleValAsnTyRAlaValAsnProLeuValSerGlyAsnGlySer 442
      .....
seq_name: pir2.G71135

seq_documentation_block:
Probable fmu protein - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
C:Accession: G71135
R:Kawarayabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino
M.A.; Ohguchi, Y.; Fuchinashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome c
A:Reference number: A71000; MUID:98344137
A:Accession: G71135
A:Status: preliminary; nucleic acid sequence not shown; transla
A:Molecule type: DNA
A:Residues: 1-450 <KAW>
A:Cross-references: GB:AP000003; NID:g3236130; PIDN:BAA29945.1;
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequen
C:Genetics:
C:Superfamily: hypotheetical protein HI0624
A:Gene: fPH0851

```

```

alignment_scores:
  Quality: 76.50      Length: 83
  Ratio: 1.628      Gaps: 3
  Percent Similarity: 56.627      Percent Identity: 27.711

alignment_block:
  US-09-049-696-7 x G71135 ..

Align seq 1/1 to: G71135 from: 1 to: 450

```

```

114  GATGCATCTGAAATTGTG.....CTGCTGAC 39
      ||||| ||||| |||||
208  AspGlyValGluValValArgSerGluArgValProThrIleLeuLysI 224
      ||||| ||||| |||||
405  GGATGGGGAAGACAAC...ACTATAAGTGGGTGCTTTAACGAG..... 79
      ||| ||||| ||||| ||||| |||||
224  eLysGlyProTyraSphenAspThrSerSerAlaPheAsnGluGlyLysI 241
      ||||| ||||| ||||| ||||| |||||
80  ....GTCAAACAAGTNGCCATCCACACAGTCGCTTTGGGGCCC 124
      ||||| ||||| ||||| ||||| |||||
241  leIleValGlnGluGluAlaSerAlaValAlaSerIleValLeuAspPro 257
      ||||| ||||| ||||| ||||| |||||
125  TCTGCAGCTCAACAACCTAGAGGAGCTCTCCAAAATGACAGAGCTTTACA 174
      ||||| ||||| ||||| ||||| |||||
258  LysProGlyGluThrValValAspLeuAlaAlaAlaProGlyGlyLysTh 274
      ||||| ||||| ||||| ||||| |||||
175  GACATATGCTTCAGATCAAGTTCAGAACATGGCCTCATTCATGCTTTT 223
      ||||| ||||| ||||| ||||| |||||
274  rThrHisLeuAlaGluLeuMetLysAsnLysGlyLysIleTyraAlaPhe 290
      ||||| ||||| ||||| ||||| |||||
seq_name: pir2.T22264

```

```
seq_documentation_block:
  C:Species: Caenorhabditis elegans
  C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change
  C:Accession: T22264
  R:Harris, B.
  submitted to the EMBL Data Library, November 1996
  A:Reference number: Z19539
```

A:Accession: T22264
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-298 <WIL>
 A:Cross-references: EMBL:Z81539; PIDN:CA804394.1; GSPDB:GN00019; CESP:F46A8.1
 A:Experimental source: clone F46A8
 C:Genetics:
 A:Gene: CESP:F46A8.1
 A:Map position: 1
 A:Introns: 30/3; 71/2; 83/1; 122/3; 202/3; 247/3; 288/2
 C:Superfamily: Caenorhabditis hypothetical protein R07C12.3

alignment_scores:
 Quality: 74.50 Length: 96
 Ratio: 1.355 Gaps: 4
 Percent Similarity: 57.292 Percent Identity: 26.042

alignment_block:
 US-09-049-696-7 x T22264 ..

Align seg 1/1 to: T22264 from: 1 to: 298

17 GGATCTGAATTTGCTGCTGACGGATGGGGAAGACAAC.....ACTAT 60
 ||||| :|||:|||||: :|||
 134 GlySerValValValIleLeuSerLysArgAsnProAsnGluValGluI 150
 61 AAGTGGGTGCTTAAACAGCTCAACAAAGTNGTGCATCATCCACACAG 110
 |||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
 150 eThrAsnGlnValAlaLysIleArgSerHisHisGlyIleValHisVal 167
 111 TCGCTTTGGG...CCCTCTGCAGCTCAAGAA.....CTAGAGGAG 148
 :||| :||| :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
 167 leAlaSerAsnThrProSerAspGlySerGlnProLeuThrLeuTyrAsp 183
 149 CTGTCCAAATGACAGAGCTTACACACATATGCTTCAGATCAAGTCA 198
 |||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
 184 LeuSerSerLysThrAsnGlyLeuThrAspPheArgAsnAspGlnPh 200
 |||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
 199 GAACAATGGCTCATTTGATGCTTTTGGG..... 226
 |||||: :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
 200 eTyrAsnSerAlaIleAspPheTyrGlyLeuAsnArgProIleAlaPheT 217
 227GCCTTCATCAGAAATGAGCTGTC 253
 :||| :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
 217 yrAlaValAsnProValValSerGlyLysGlySerIle 229

seq_name: pirl:NPKEDC

seq_documentation_block:
 anthranilate phosphoribosyltransferase (EC 2.4.2.18) - Acinetobacter calcoaceticus
 C:Species: Acinetobacter calcoaceticus
 C:Date: 28-May-1986 #sequence_revision 28-May-1986 #text_change 11-Jun-1999
 C:Accession: A00585
 R:Kaplan, J.B.; Goncharoff, P.; Seibold, A.M.; Nichols, B.P.
 Mol. Biol. Evol. 1, 456-472, 1984
 A:Title: Nucleotide sequence of the Acinetobacter calcoaceticus trpGDC gene cluster.
 A:Reference number: A93052; MUID:88174326
 A:Accession: A00585
 A:Molecule type: DNA
 A:Residues: 1-349 <RAP>
 A:Cross-references: GB:M36636; NID:g141799; PIDN:AAA21904.1; PID:g141801
 C:Comment: This enzyme catalyzes the formation of anthranilate by transferring the phosphate carboxyl-terminal two-thirds of component II of anthranilate synthase (EC 4.1.3.27).
 C:Genetics:
 A:Gene: trpD
 C:Superfamily: anthranilate phosphoribosyltransferase; trpD homology
 C:Keywords: glycosyltransferase; pentosyltransferase; tryptophan biosynthesis
 F:6-339/Domain: trpD homology <TRD>

alignment_scores:
 Quality: 74.00 Length: 57

Ratio: 1.947 Gaps: 0
 Percent Similarity: 66.667 Percent Identity: 28.070
 alignment_block:
 US-09-049-696-7 x NPKEDC ..
 Align seg 1/1 to: NPKEDC from: 1 to: 349
 20 TCTGAATTTGCTGCTGACGGATGGGGAAGACAACACTATAAGTGGGTG 69
 :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
 293-AlaAsnMetIleAlaLeuAsnAlaGlyAlaGlyIleTyrValSerGlyLe 309
 70 CTTTAAACGAGGTCAACAAAGTNGTGCATCATCCACACAGTCGCTTTGG 119
 :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
 309 uAlaThrSerTyrLysGlnGlyValAlaLeuAlaHisAspIleIleTyrG 326
 120 GGCCTCTGCAGCTCAAGAACTAGAGGAGCTGCCAAAATGACAGGAGT 169
 || :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
 326 yGlyGlnAlaLeuGluLysMetSerIleLeuSerGluPheThrLysAla 342
 170 TTACACACATATGCTTCAGAT 190
 \ |||||: |||||:|||||: :|||:|||||: :|||:|||||:
 343 LeuLysGluTyrAlaAsnAsn 349

seq_name: pirl:T15069

seq_documentation_block:
 hypothetical protein R07C12.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 10-Dec-1999
 C:Accession: T15069
 R:Geisel, C.; Beck, C.; O'Brien, D.
 submitted to the EMBL Data Library, December 1997
 A:Description: The sequence of C. elegans cosmid R07C12.
 A:Reference number: Z18286
 A:Accession: T15069
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-358 <GET>
 A:Cross-references: EMBL:AF038622; NID:g2736331; PID:g2736334; PIDN:AAB94152.1; GSPDB
 A:Experimental source: strain Bristol N2; clone R07C12
 C:Genetics:
 A:Gene: CESP:R07C12.3
 A:Map position: 4
 A:Introns: 9/3; 86/2; 117/3; 197/3; 247/1; 292/3; 333/2
 C:Superfamily: Caenorhabditis hypothetical protein R07C12.3

alignment_scores:
 Quality: 72.50 Length: 97
 Ratio: 1.368 Gaps: 4
 Percent Similarity: 54.639 Percent Identity: 26.804
 alignment_block:
 US-09-049-696-7 x T15069 ..
 Align seg 1/1 to: T15069 from: 1 to: 358

8 CCAACTGATGATCTGAAATTTGCTGCTGACG.....CATGGGAAGA 51
 ||| ||||| :|||:|||||: :|||:|||||: :|||:|||||:
 126 ProLeuCysGlySerIleValIleValIleLeuSerLysArgAsnProAsnG 142
 52 CAACACTATAAGTGGGTGCTTTTACGAGGTCAACAAAGTNGTCCCATCA 101
 :||| ||||| :|||:|||||: :|||:|||||: :|||:|||||:
 142 uAsnAspIleSerLysLeuValAlaLysIleArgSerHisHisGlyIleV 159
 102 TCCACACAGTCGCT.....TTGGGCGCTCTGCAGCTCAAGAA 139
 :|||:|||||: :|||:|||||: :|||:|||||: :|||:|||||:
 159 aHisValIleAlaSerAsnThrProSerGlyGlySerGlnProLeuThr 175
 140 CTAGAGGAGCTGTCCAAAATGACAGGAGGTTCACAGACATATGCTTCAGA 189
 ||| :|||:|||||: :|||:|||||: :|||:|||||:
 176 LeuTyrAspLeuSerSerArgThrAsnGlyLeuThrAspPheArgAsnAs 192

190 T.....CAAGTTCAGACAATGGCTC..... 211
192 pAspGlnLeuAlaLeuSerArgLeuProThrValGlyLeuTyrTrpProL 209
212ATTGATGCTTTGGGGCCCTTCATCAGGAATGGA 247
209 euProLeuTyrAlaGluAsnSerIleValSerGlyAsnGly 222

seq_name: pir2:F84811

seq_documentation_block:

probable retroelement pol polyprotein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: F84811

M.: Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, L.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: F84811

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-689 <STO>

A:Cross-references: GB:AE002093; NID:93928084; PIDN:AACT9610.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g38970

A:Map position: 2

alignment_scores:

Quality: 71.00 Length: 108

Ratio: 1.092 Gaps: 5

Percent Similarity: 60.185 Percent Identity: 24.074

alignment_block:

US-09-049-696-7 x F84811 ..

Align seg 1/1 to: F84811 from: 1 to: 689

20 TCTGAAATGTGCTGCTGACGATGGGAGACAAAC..... 55
348 AlaserlleleuLeuSerAspGlyArgAspThrTyrMetAsnG1 364

56ACTATAAGTGGTGCT 71

364 nAlaAspProAsnTyrLysLeuLeuLeuProLeuSerMetHisGlyCys. 380

72 TTAACGAGGTCAACAAGTNGTCCATCATCCACAGTCGCTTTGGG 121

381GluSerLysArgPheGlnIleProValHisSerPheGlyPheGly 395

122 CCCTCT...GCAGCTCAAGAACTAGAGAGCTGTCCAAATGACAGGAGG 168

396 SerAspHisAspAlaSerLeuMetHisSerValSerGluThrSerGlyG1 412

169 TTTACAGACATAT...GCTTCAGATCAAGTTTCAGAACAAATGCCTCATPTG 215

412 yThrPheSerPheIleGluSerGluSerValIleGlnAspAlaLeuAlaG 429

216 ATGCTTTTGGGGCCCTTCATCAGGAATGGAGTGCTCTCAGCGGCTCC 265

429 IncysIleGlyGlyLeuLeuSer.....ValAlaValGlnGlu 441

266 ATCCAGCTTGAGTAAGGGATTA 289

442 LeuArgLeuGluIleGluGlyMet 449

seq_name: pir1:PWPA

seq_documentation_block:

H+-transporting ATP synthase (EC 3.6.1.34) alpha chain - radish mitochondrion

H+-transporting ATP synthase (EC 3.6.1.34) alpha chain - rape mitochondrion
C:Species: mitochondrion Brassica napus (rape)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C:Accession: S13382; S30090
R:Handa, H.; Nakajima, K.
Plant Mol. Biol. 16, 361-364, 1991
A:Title: Nucleotide sequence and transcription analyses of the rapeseed (Brassica napus) H+-transporting ATP synthase alpha chain
A:Reference number: S13382; MUID:91370881
A:Accession: S13382
A:Molecule type: DNA
A:Residues: 1-507 <HAN>

A:Cross-references: EMBL:X56008; NID:919553; PIDN:CAA39483.1; PID:919554
A:Experimental source: strain Polima
R:Bonhomme, S.; Budar, F.; Lancelin, D.; Small, I.; Defrance, M.C.; Pelletier, G.
Mol. Gen. Genet. 235, 340-348, 1992

A:Title: Sequence and transcript analysis of the Nco2.5 Ogura-specific fragment corresponding to the H+-transporting ATP synthase alpha chain
A:Reference number: S30089; MUID:93101139
A:Accession: S30090
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-111 <BON>

A:Cross-references: EMBL:Z12627; NID:914388; PIDN:CAA78274.1; PID:914389

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992

C:Genetics:

A:Gene: atpA

A:Genome: mitochondrion

C:Superfamily: H+-transporting ATP synthase alpha chain; H+-transporting ATP synthase
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion;
F:171-178/Region: nucleotide-binding motif A (P-loop)

F:205-377/Domain: H+-transporting ATP synthase alpha chain homology <NTP>

F:268-273/Region: nucleotide-binding motif B

F:177/Binding site: ATP (Lys) #status Predicted

F:376/Active site: Arg #status Predicted

alignment_scores:

Quality: 70.50 Length: 92

Ratio: 1.259 Gaps: 2

Percent Similarity: 60.870 Percent Identity: 25.000

alignment_block:

US-09-049-696-7 x PWPA ..

Align seg 1/1 to: PWPA from: 1 to: 507

23 GAAATTCGTGCTGACGATGGGAGACACACATATAAGTGGTGCTT 72

53 GluMetValLeuPheAlaAsnGlyValLysGlyMetAlaLeuAsnLeuG1 69

73 TAACGAGGTCAACAAGTNGTCCATCATCCACAGTCGCTTTGGGC 122

69 uasnGlu.....AsnValGlyIleValPheGlyG 80

123 CCTCTGAGCTCAAGAACTAGAGGAGCTGTCCAAATGACAGGAGTTTA 172

80 lyAspThrAlaIleLysGluLysLeuValLysArgThrGlySerIle 96

173 CAGACATATGCTTCAGATCAAGTTTCAGAACAAATGCCTCATTTGCTTT 222

97 ValAspValProAlaGlyLysAlaMetLeuGlyArgValValAspAlaMe 113

223 TGGGGCCCTTCATCAGGAATGGAGTGCTCT.....CAGCGCT 263

113 tGlyValProIleAspGlyArgGlyAlaLeuSerAspHisGluGlnArgA 130

264 CCATCCAGCTTGAGTAAGGGATTA 289

135 rgValGluValLysAlaProGlyIle 138

seq_name: pir2:S12309

seq_documentation_block:

H+-transporting ATP synthase (EC 3.6.1.34) alpha chain - radish mitochondrion

C:Species: mitochondrion Raphanus sativus (radish)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C:Accession: S12309
R:Makarov, C.A.; Apel, I.J.; Palmer, J.D.
Plant Mol. Biol. 15, 735-746, 1990
A:Title: Characterization of radish mitochondrial atpA: influence of nuclear background
F:171-178/Region: nucleotide-binding motif A (P-loop)
F:205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

alignment_scores:
Quality: 70.50 Length: 100
Ratio: 1.306 Gaps: 3
Percent Similarity: 54.000 Percent Identity: 24.000

alignment_block:
US-09-049-696-7 x T15070
Align seg 1/1 to: T15070 from: 1 to: 516

2 AAATATCAACTGATGATCTGAATATGCTGCTGACGATGGGAAGA 51
314 GlnPheSerThrCysGlySerLleValLeuSerLysArgAsnPr 330
52 CAAC.....ACTATAAGTGGTGCTTTACGAGGTCACAAAGTNGTG 95
330 oAsnGluValGluIleThrAsnLeuValAlaLysIleArgSerHisG 347
96 CCATCATCCACACAGTCGCT.....TTGGGGCCCTCTGCAGCT 133
347 lylleValHisValIleSerSerAsnThrProSerGlySerGlnPro 363
134 CAAGAACTAGAGAGCTGTCCAAATGACAGGAGGTTTACAGCATATGC 183
364 LeuThrLeuTyAspLeuSerSerLysThrAsnGlyLeuAlaAspPheAr 380
184 TTCAGATCAAGTTCAGACAAATGCCCTCATTGAT..... 217
380 gAsnAspAspGluPheAspGluSerValIleProPheAlaGlyLeuAsnT 397
218CCTTTTGGGGCCCTTTCATCAGGAAATCGAGCT 250
397 yrProIleValAsnTyrrAlaValAsnProLeuValSerGlyAsnGlySer 413

seq_name: pir2:F96584

seq_documentation_block:
hypothetical protein F20D21.13 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: F96584
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: F96584
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-260 <STO>
A:Cross-references: GB:AE005173; NID:g4585975; PIDN:AAD25611.1; GSPDB:GN00141
C:Genetics:
A:Gene: F20D21.13
A:Map position: 1

alignment_scores:
Quality: 69.50 Length: 126
Ratio: 1.086 Gaps: 5
Percent Similarity: 50.794 Percent Identity: 22.222

alignment_block:
US-09-049-696-7 x F96584

C:Species: mitochondrion Raphanus sativus (radish)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C:Accession: S12309
R:Makarov, C.A.; Apel, I.J.; Palmer, J.D.
Plant Mol. Biol. 15, 735-746, 1990
A:Title: Characterization of radish mitochondrial atpA: influence of nuclear background
F:171-178/Region: nucleotide-binding motif A (P-loop)
F:205-377/Domain: H+-transporting ATP synthase alpha chain homology <ATP>

alignment_scores:
Quality: 70.50 Length: 92
Ratio: 1.259 Gaps: 2
Percent Similarity: 60.870 Percent Identity: 25.000

alignment_block:
US-09-049-696-7 x S12309
Align seg 1/1 to: S12309 from: 1 to: 507

23 GAAATGCTGCTGCTGCGGATGGGAGACACACTATATAGTGGTGCT 72
53 GluMetValLeuPheAlaAsnGlyValLysGlyMetAlaLeuAsnLeuG 69
73 TAAGAGGTCACAAAGTNGTCCATCATCCACAGTCGCTTTGGGGC 122
69 uAsnGlu.....AsnValGlyIleValPheGlyG 80
123 CCTCTGCGAGCTCAAGAACTAGAGGAGCTGTCCAAATGACAGGAGTTTA 172
80 lyAspThrAlaIleLysGluGlyAspLeuValLysArgThrGlySerIle 96
173 CAGACATATGCTTCAGATCAAGTTCAGACAAATGGCTCTATTGATGCTT 222
97 ValAspValProAlaGlyLysAlaMetLeuGlyArgValValAspAlaMe 113
223 TGGGGCCCTTTCATCAGGAAATGAGCTGCTCT.....CAGCGCT 263
113 tGlyValProIleAspGlyArgGlyAlaLeuSerAspHisGluGlnArgA 130
264 CCATCCAGCTTGAGAGTAAGGATTA 289
130 rgValGluValLysAlaProGlyIle 138

seq_name: pir2:T15070

seq_documentation_block:
hypothetical protein R07C12.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15070
R:Geisdel, C.; Beck, C.; O'Brien, D.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid R07C12.
A:Reference number: Z18286
A:Accession: T15070
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-516 <GEL>
A:Cross-references: EMBL:AF038622; NID:g2736331; PID:g2736333; PIDN:AAB94151.1; GSPDB:GN
A:Experimental source: strain Bristol N2; clone R07C12
C:Genetics:
A:Gene: CESP:R07C12.2

Align seg 1/1 to: F96584 from: 1 to: 260

```

2 AAATATCAACTGATGATGTAATTTGCTGCTGACGGATGGGGAAGA 51
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : : :
80 LysProProGlnThrGlyAspValValIleValAlaAspGlyThrG1 96
   : : : : : : : : : : : : : : : : : : : : : : : : : :
52 CAACACTATAGT...GGTCTTTTACGAGGTCAA...C 86
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : :
96 uasProIleGlyTrpGlyLeuTyAsnSerValSerMetPheCysValA 113
   : : : : : : : : : : : : : : : : : : : : : : : : : :
87 AAAGTNGTGCATCATCACACACAGCTGCTTTGGGGCCCTCTGCAGCTCAA 136
   : : : : : : : : : : : : : : : : : : : : : : : : : :
113 rgLeuMetGlnLeuGlnHisGluSerThrArgAspProSerCysAlaLeu 129
   : : : : : : : : : : : : : : : : : : : : : : : : : :
137 GAACATAGAGAGCTGTCCAAATGACAGGAGTTTACAGACATATGCTTC 186
   : : : : : : : : : : : : : : : : : : : : : : : : : :
130 AsnIleGluLysLeu.....LeuGlnThrArgIleAl 140
187 AGATCAAGTTTCAG..... 199
140 aGluAlaValGlnLeuArgLysSerLeuAlaLeuProSerAlaAsnThrA 157
200 .....AACAAATGGGCTC 211
157 snAlaTyArgLeuValAsnSerGluGlyAspArgLeuSerGlyLeuIle 173
212 ATTGATGCTTTTGGG.....GCCCTTTCATCAGGAAATGAGAGTGT 252
   : : : : : : : : : : : : : : : : : : : : : : : : : :
174 ValAspValPheGlyAspIleAlaValValAlaSerSerAlaAlaTrpLe 190
253 CTCTCAGCGCTCCATCCAGCTTGAGAGT 280
190 uGluLysTyArgIleGluValGluSer 199

```

seq_name: pir2:T37223

seq_documentation_block:

probable secreted protein - Streptomyces coelicolor

C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C;Accession: T37223

R;Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998

A;Reference number: Z21615

A;Accession: T37223

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-421 <OLI>

A;Cross-references: EMBL:AL031514; PIDN:CAA20601.1; GSPDB:GN00070; SCOEDB:SC2H4.07c

A;Experimental source: strain A3(2)

C;Genetics:

A;Gene: SCOEDB:SC2H4.07c

alignment_scores:

Quality	69.50	Length:	85
Ratio:	1.337	Gaps:	8
Percent Similarity:	61.176	Percent Identity:	35.294

alignment_block:

US-09-049-696-7 x T37223 ..

Align seg 1/1 to: T37223 from: 1 to: 421

```

14 GATGGATCTGAA...ATTGTCTGCTGCTGAGGAGGGAAGACAAC..... 55
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : :
147 AspGlySerLysArgIleValLeuIleSerAspGlyGluAspThrCysAl 163
   : : : : : : : : : : : : : : : : : : : : : : : : : :
56 .ACTATAAGTGGGTGCTTTTACAGAGGTCAACAAAGTNGTGCC..... 97
   : : : : : : : : : : : : : : : : : : : : : : : : : :
163 aProLeuAspProCys.....GluValAlaArgGluIleAlaAlaLysG 178

```

```

98 .....ATCATCCACACAGTCGCTTTGGGGCCCTCTGCA..... 130
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : :
173 lylleGlyLeuThrIleAspThrLeuGlyLeuValProAsnThrLysMet 194
   : : : : : : : : : : : : : : : : : : : : : : : : : :
131 GCTCAAGAAGTAGAGGAGCTGTCCAAAATGACAGGAGGTTTACAGACATA 180
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : :
195 ArgGlnGlnLeuSerCysIleAlaGluAlaThrGlyGly.....ThrTy 209
   : : : : : : : : : : : : : : : : : : : : : : : : : :
181 TGCTTCA.....GATCAAGTTTCAGAAC.....AATGGCCTCA 212
   ||| ||| : : : : : : : : : : : : : : : : : : : : : : :
209 rThrSerValGluHisThrAspGluLeuThrAspLysValAsnGlnLeuV 226
213 TTGAT 217
226 alasp 227

```

This Page Blank (uspto)

Sequence	Strid Orig	zScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-477-396A-2	+	70.50	132.78	0.7823
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-477-396A-2	+	68.50	141.23	0.9695
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-09-1273-492A-1	+	66.50	137.75	1.70
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-426-627-4	+	66.50	123.47	2.57
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-426-627-24	+	66.50	123.47	2.57
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-426-627-6	+	66.50	122.76	2.63
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-426-627-23	+	66.50	122.75	2.63
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-09-061-400-2	+	65.00	113.82	4.81
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-09-001-273-2	+	65.00	113.71	4.93
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-843-459A-2	+	65.00	113.71	4.83
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-387-942C-46	+	64.50	135.37	2.89
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-387-942C-4	+	64.50	116.34	5.02
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-772-270A-11	+	62.50	111.19	9.23
/cgn2_6/ptodata/2/iaa/PCTUS.COMB	pep:PCT-US93-10500-2	+	62.50	109.47	9.71
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-305-223-468	+	62.00	132.68	5.55
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-246-982A-6	+	62.00	98.97	14.78
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-453-265-6	+	62.00	98.97	14.78
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-457-273B-42	+	62.00	98.97	14.78
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-556-419-21	+	62.00	98.97	14.78
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-09-041-886-15	+	62.00	98.97	14.78
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-09-188-930-125	+	61.50	120.53	8.86
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-09-380-930-2	+	61.50	114.93	10.43
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-08-924-345-3	+	61.50	114.22	10.64
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-191-866B-58	+	61.50	114.22	10.64
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-185-949B-58	+	61.50	114.22	10.64
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-09-188-930-303	+	61.50	114.22	10.64
/cgn2_6/ptodata/2/iaa/6A.COMB	pep:US-09-310-293-2	+	61.50	107.28	13.02
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-508-448C-19	+	61.00	122.92	9.27
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-508-448C-25	+	61.00	116.99	11.02
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-742-273-2	+	61.00	114.98	11.68
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-387-942C-51	+	60.50	125.96	9.53
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-387-942C-5	+	60.50	108.41	15.86
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-08-818-112-114	+	60.00	115.95	14.29
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-231-342-6	+	60.00	115.90	14.32
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-09-055-957-1	+	60.00	115.68	14.41
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-687-56A-23	+	60.00	101.35	21.85
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-418-8930-23	+	60.00	98.45	23.77
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-418-8930-24	+	60.00	98.45	23.77
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-08-961-083-96	+	59.50	119.27	14.56
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-102-863-11	+	59.50	113.65	17.15
/cgn2_6/ptodata/2/iaa/PCTUS.COMB	pep:PCT-US92-10885-11	+	59.50	113.65	17.15

331 GUUGLYASHTNIR IIEGIUIIEGLYCYSASPGLY

64 TGGGTGC.....TTTAACGAGGTCAAACAAGTNGTGCCATCA 101

seq name: /cqn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-477-396A-2

23 GAAATTGTGCTGCTGACGGATGGG.....GA 48

alignment_scores:		
Quality:	66.50	Length: 85
Ratio:	1.357	Gaps: 3
Percent Similarity:	57.647	Percent Identity: 25.882

```
alignment_block:
US-09-049-696-7 x US-09-123-492A-1 ..
Align^seg 1/1 to: US-09-123-492A-1 from: 1 to: 189
```

23 GAAATTGTGCTGCTGACGGATGGG.....GA 48

68 AspValValValLeuProGlyGlyAsnLeuGlyAlaGlnAsnLeuSerGI 84
49 AGACAACTATAAGTGGTCTTTAAACGAGTCAACAAAGTNGTCCA 98
84 uSerAlaAlaValLysGluIleLeuLysGluGlnGluAsnArgLysGlyL 101
99 TCATCCACAGAGTGGCTTTGGGGCCCTCTGCA.....GCTCAAGAACTA 142
101 euIleAlaAlaIleCysAlaGlyProThrAlaLeuLeuAlaHisGluIle 117
143 GAGGAGCTGTCCAAATGACA..... 163
118 GlyCysGlySerLysValThrHisProLeuAlaLysAspLysMetMe 134
164GGAGTTTACAGATCATGTTTCAGATCAAGTTCAGAACATGGCC 209
134 tAsnGlyGlyHisThrThrTyrSerGluAsnArgValGluLysAspGlyL 151
210 TCATT 214
151 euIle 152

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.US-08-426-627-4

seq_documentation_block:
Sequence 4, Application US/08426627
Patent No. 5756664
GENERAL INFORMATION:
APPLICANT: Amann, Egon
APPLICANT: Otawara-Hamamoto, Yoko
APPLICANT: Kikuno, Reiko
APPLICANT: Takeshita, Sunao
APPLICANT: Tezuka, Kenichi
TITLE OF INVENTION: No. 5756664el Protein with Bone Formation
TITLE OF INVENTION: Ability and Process for Its Production.
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/426,627
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/036,841
FILING DATE: 25-MAR-1993
APPLICATION NUMBER: JP 4-71501
FILING DATE: 27-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hammond, Alan W.
REGISTRATION NUMBER: 35,178
REFERENCE/DOCKET NUMBER: 02481-1285-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 779 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-426-627-4

alignment_scores:
Quality: 66.50 Length: 99
Ratio: 1.167 Gaps: 4
Percent Similarity: 57.576 Percent Identity: 25.253
alignment_block:
US-09-049-696-7 x US-08-426-627-4 ..
Align seg 1/1 to: US-08-426-627-4 from: 1 to: 779
14 GATGATCTGAAATTGCTGCTGAGGATGGGAGAACACACTATAG 63
325 GluGlyAsnThrIleGluIleGlyCysAspGlyAspSerIleThrValAs 341
54 TGGGTGC.....TTTAACGAGGTCAACAAAGTNGTCCATCA 101
341 nGlyLysMetValAsnLysLysAspIleValThrAsnAsnGlyValI 358
102 TCCACACAGTC.....GCTTTGGGCCCTCTGCAGCTCAAGAACTAGAG 145
358 leHisLeuIleAspGlnValLeuIleProAspSerAlaLysGlnValIle 374
146 GAGCTGTCCAAATGACAGGAGGTTTACAGACATATGCTTCAGATCAAGT 195
375 GluLeuAla.....GlyLysGlnGlnThrThrPheThrAspLeuVa 388
196 TCAGAACAAATGCCCTCATTTGATGCT.....TTTGGGG 227
388 lalaglnLeuGlyLeuAlaSerAlaLeuArgProAspGlyGlyThrL 405
228 CCCTTTTCATCAGAAATGAGCTGTCTCAGCGCTCCATCCAGCTT 274
405 euLeuAlaProValAsnAsnAlaPheSerAspThrLeuSerMet 420
seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.US-08-426-627-4

seq_documentation_block:
Sequence 24, Application US/08426627
Patent No. 5756664
GENERAL INFORMATION:
APPLICANT: Amann, Egon
APPLICANT: Otawara-Hamamoto, Yoko
APPLICANT: Kikuno, Reiko
APPLICANT: Takeshita, Sunao
APPLICANT: Tezuka, Kenichi
TITLE OF INVENTION: No. 5756664el Protein with Bone Formation
TITLE OF INVENTION: Ability and Process for Its Production.
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/426,627
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/036,841
FILING DATE: 25-MAR-1993
APPLICATION NUMBER: JP 4-71501
FILING DATE: 27-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hammond, Alan W.
REGISTRATION NUMBER: 35,178
REFERENCE/DOCKET NUMBER: 02481-1285-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 779 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

[illegible]

```

: TITLE OF INVENTION: No. 5756664el Protein with Bone Formation
: TITLE OF INVENTION: Ability and Process for its Production.
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
: ADDRESSEE: Dunner
: STREET: 1300 I Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/426,627
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/036,841
: FILING DATE: 25-MAR-1993
: APPLICATION NUMBER: JP 4-71501
: FILING DATE: 27-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Hammond, Alan W.
: REGISTRATION NUMBER: 35,178
: REFERENCE/DOCKET NUMBER: 02481-1285-00000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-408-4000
: TELEFAX: 202-408-4400
: INFORMATION FOR SEQ ID NO: 23:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 837 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-426-627-23

```

```

alignment_scores:
  Quality: 66.50      Length: 99
  Ratio: 1.167       Gaps: 4
Percent Similarity: 57.576 Percent Identity: 25.253

```

```
alignment_block:
US-09-049-696-7 x US-08-426-627-23

```

```
Align seg 1/1 to: US-08-426-627-23 from: 1 to: 837
```

```

14 GATGATCTGAATGTGCTGCTGCTGACGGATGGGGAAGACACACTATTAAG 63
   ::::: ||| ::| |||||: ||:::
325 GluGlyAsnThrIleGluIleGlyCysAspGlyAspSerIleThrValas 341
   ::::: ||| ::| |||||: ||:::
64 TGGGTGC.....TTTACGAGGTCAAAACAAAGTNGCCATCA 101
   ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
341 nGlyIleLysMetValAsnLysLysAspIleValThrAsnAsnGlyVali 358
   ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
102 TCCACACAGTC.....GCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAG 145
   ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
358 leHisLeuIleAspGlnValLeuIleProAspSerAlaLysGlnValile 374
   ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
146 GAGCTGTCCAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAGT 195
   |||||: ||| ::| ::| ::| ::| ::| ::| ::|
375 GluLeuAla.....GlyLysGlnGlnThrThrPheThrAspLeuVa 388
   ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
196 TCAGAACAAATGGCTCATTTGATGCT.....TTTGGG 227
   ||| ::| ||||| ::||| ::|||
388 lAlaGlnLeuGlyLeuAlaSerAlaLeuArgProAspGlyGluThrL 405
   |||||: ||| ::| ::| ::| ::| ::| ::| ::|
228 CCCTTTTCATCAGGAATGGAGCTCTCTCAGCGCTCCATCAGCTT 274
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||:::

```

```

: 405 euLeuAlaProValAsnAsnAlaPheSerAspThrLeuSerMet 420
: seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-09-061-400-2
: seq_documentation_block:
: Sequence 2: Application US/09061400
: Patent No. 6077936
: GENERAL INFORMATION:
: APPLICANT: SHYJAN, Andrew
: TITLE OF INVENTION: NOVEL MULTIDRUG RESISTANCE-ASSOCIATED
: TITLE OF INVENTION: POLYPEPTIDE
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD, LLP
: STREET: 28 State Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/061,400
: FILING DATE: 16-APRIL-1998
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Elizabeth A. Hanley
: REGISTRATION NUMBER: 33,505
: REFERENCE/DOCKET NUMBER: MNI-056CP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 227-7400
: TELEFAX: (617) 742-4214
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1437 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-061-400-2

```

```

alignment_scores:
  Quality: 65.00      Length: 98
  Ratio: 1.083       Gaps: 3
Percent Similarity: 61.224 Percent Identity: 25.510

```

```
alignment_block:
US-09-049-696-7 x US-09-061-400-2

```

```
Align seg 1/1 to: US-09-061-400-2 from: 1 to: 1437
```

```

2 AAATATCCAACGTGATGGATCTCAAAATTGCTGCTGACGGATGGG..... 46
   ::||| ::||| ::|||: |||||: ||:::
745 GlnThrLeuValAspCysAspGluValIlePheMetLysGluGlyCysl 761
   ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
47 ....GAAGACACACTATAATAGTGGTGGTCTTTAACGAGGTCAACAAAGTN 92
   ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
761 ethrGluArgGlyThrHisGluLuleuMetAsnLeuAsnGlyAspTyrA 778
   ::|||: ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
93 GTGCCATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGACTA 142
   ::|||: ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
778 lThrIlePheAsnAsnLeuLeuLeuGlyGluThrProValGluIle 794
   ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
143 GAGGAGCTGTCCAAATGACAGAGGTTTACAGACATATGCTTCAGATCA 192
   ::|||: ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
795 AsnSer...LysLysGluThrSerGlySerGlnLysLysSerGlnAspLy 810
   |||||: ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
193 AGTTCAGAACAAATGGCTTCATGATGCTTTTGGGGCCCTTTTCATCAGGAA 242
   ::|||: ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
870 sGlyProLysThrGlySerIleLysLysLysAlaValLysProGluG 827

```

us-09-049-696-7.ra1

Tue Apr 2 09:40:12 2002

```

193 AGTTCAGAACATGGCTCATTGCTTTTGGGCCCTTTTCATCAGGAA 242
      ::::::::::: |||
826 SGlyProLysThrGlySerIleLysLysGluLysAlaValLysProGluG 843
      ::::::::::: |||
243 ATGAGAGTGTCTCTCAGCGCTCCATCCAGCTTCAGAGTAAGGA 286
      ::::::::::: |||
843 luGlyGlnLeu.....ValGlnLeuGluGluLysGly 853
      ::::::::::: |||

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pap:US-08-843-459A-2
seq_documentation_block:
; Sequence 2, Application US/08843459A
; Patent No. 6162616
; GENERAL INFORMATION:
; APPLICANT: SHYJAN, Andrew
; TITLE OF INVENTION: NOVEL MULTIDRUG RESISTANCE-ASSOCIATED
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/843,459A
; FILING DATE: 16-APR-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanley, Elizabeth A.
; REGISTRATION NUMBER: 33,505
; REFERENCE/DOCKET NUMBER: MNI-056 (formerly MIL-001)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1453 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-843-459A-2

alignment_scores:
  Quality: 65.00      Length: 98
  Ratio: 1.083        Gaps: 3
  Percent Similarity: 61.224      Percent Identity: 25.510

alignment_block:
US-09-049-696-7 x US-08-843-459A-2 ..
Align seg 1/1 to: US-08-843-459A-2 from: 1 to: 1453
      2 AAATATCAACTGATGATCTGAATTTGCTGCTGACGATGG..... 46
      761 GlnTyrLeuValAspCysAspGluValIlePheMetLysGluGlyCysI 777
      47 ....GAAGACAACACTATAAGTGGTCTTTAACGAGGTCAACAAAGTN 92
      777 eThrGluArgGlyThrHisGluGluLeuMetAsnLeuAsnGlyAspTyrA 794
      93 GTGCATCATCCACACAGCTGCTTTGGGCCCTCTGCAGCTCAAGAACTA 142
      794 laThrIlePheAsnAsnLeuLeuGluGluThrProProValGluIle 810
      143 GAGGAGCTGTCCAAATGACAGGAGGTTCACAGCATATGCTTCAGATCA 192

```

```

243 ATGAGAGTGTCTCTCAGCGCTCCATCCAGCTTCAGAGTAAGGA 286
      ::::::::::: |||
827 luGlyGlnLeu.....ValGlnLeuGluGluLysGly 837
      ::::::::::: |||

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pap:US-09-001-273-2
seq_documentation_block:
; Sequence 2, Application US/09001273
; Patent No. 5994130
; GENERAL INFORMATION:
; APPLICANT: SHYJAN, Andrew
; TITLE OF INVENTION: NOVEL MULTIDRUG RESISTANCE-ASSOCIATED
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Testa, Hurwitz & Thibault
; STREET: 125 High St.
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/001,273
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: FENTON, Gillian M
; REGISTRATION NUMBER: 36,508
; REFERENCE/DOCKET NUMBER: MIL-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1453 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-001-273-2

alignment_scores:
  Quality: 65.00      Length: 98
  Ratio: 1.083        Gaps: 3
  Percent Similarity: 61.224      Percent Identity: 25.510

alignment_block:
US-09-049-696-7 x US-09-001-273-2 ..
Align seg 1/1 to: US-09-001-273-2 from: 1 to: 1453
      2 AAATATCAACTGATGATCTGAATTTGCTGCTGACGATGG..... 46
      761 GlnTyrLeuValAspCysAspGluValIlePheMetLysGluGlyCysI 777
      47 ....GAAGACAACACTATAAGTGGTCTTTAACGAGGTCAACAAAGTN 92
      777 eThrGluArgGlyThrHisGluGluLeuMetAsnLeuAsnGlyAspTyrA 794
      93 GTGCATCATCCACACAGCTGCTTTGGGCCCTCTGCAGCTCAAGAACTA 142
      794 laThrIlePheAsnAsnLeuLeuGluGluThrProProValGluIle 810
      143 GAGGAGCTGTCCAAATGACAGGAGGTTCACAGCATATGCTTCAGATCA 192
      811 AsnSer...LysLysGluThrSerGlnLysLysSerGlnAspLys 826

```



```

111 AnSer...LysLysGluThrSerGlySerGlnLysLysSerGlnAspGly 826
193 AGTTGAGAACAAATGCGCTCATTTGATGCTTTTGGGGCCCTTTTCATCAGGAA 242
826 sGlyProLysThrGlySerLysLysGluLysAlaValLysProGluG 843
243 ATGAGAGTGTCTCAGCGGTCCTCCAGCTTGCAGCTTGCAGTAAGGGA 286
843 LuGlyGlnLeu.....ValGlnLeuGluGluLysGly 853
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-387-942C-46

```

```

seq_documentation_block:
Sequence 46, Application US/08387942C
Patent No. 5939289
GENERAL INFORMATION:
APPLICANT: ERTESVAG, HELGA
APPLICANT: VALLA, SVEIN
APPLICANT: SKJAK-BRAEK, GUDMUND
TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES
TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
STREET: P.O. BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22042
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/387,942C
FILING DATE: 09-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR, GERALD M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1809-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8050
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-387-942C-46

```

```

alignment_scores:
Quality: 64.50 Length: 92
Ratio: 1.217 Gaps: 2
Percent Similarity: 57.609 Percent Identity: 26.087

```

```

alignment_block:
US-09-049-696-7 x US-08-387-942C-46
Align seg 1/1 to: US-08-387-942C-46 from: 1 to: 151

```

```

11 ACTGATGATCTGAATTTGCTGCTGACGGTGGGAGACACATAT 60
:|||||: |||:|||||:
13 SerAspAlaHisGluThrLeuGlyLeuAspGlyAsnArgLeuAs 29
:|||||: |||:|||||:
61 AAGTGGGTGCTTTAACGAGTCAACAAAGTNGTCCCATCACCACAG 110
:|||||: |||:|||||:

```

```

29 nGlyGlyAlaGlyAsnAspIleLeuAspGlyGlyAlaGlyArgAspAsnL 46
111 TCGTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGGAGCTGCCAAATG 160
:|||||: |||:|||||:
46 euThrGlyGlyAlaGlyAla.....AspLeuPheArgVal 57
161 ACAGAGGTTTACAGACATATGCTTCAGATCAAGTTTCAGAACAAATGGCCT 210
:|||||: |||:|||||:
58 SerAlaArgThrAspSerTyrArgThrAspSerAlaSerPheAsnAspLe 74
211 CATTGATGCTTTTGGGGCCCTTTCATCAGGAATGAGGCTGCTCTCAGC 260
:|||||: |||:|||||:
74 uilleThrAspPheAspAla.....SerGlnA 83
261 GCTCATCCAGCTTGAGAGTAAGGA 286
:|||||: |||:
83 spArgIleAspLeuSerAlaLeuGly 91
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-387-942C-4
seq_documentation_block:
Sequence 4, Application US/08387942C
Patent No. 5939289
GENERAL INFORMATION:
APPLICANT: ERTESVAG, HELGA
APPLICANT: VALLA, SVEIN
APPLICANT: SKJAK-BRAEK, GUDMUND
TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES
TITLE OF INVENTION: ENCODING MANNURONAN C-5-EPIMERASE
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
STREET: P.O. BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22042
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/387,942C
FILING DATE: 09-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR, GERALD M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1809-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 997 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-387-942C-4
alignment_scores:
Quality: 64.50 Length: 92
Ratio: 1.217 Gaps: 2
Percent Similarity: 57.609 Percent Identity: 26.087
alignment_block:
US-09-049-696-7 x US-08-387-942C-4
Align seg 1/1 to: US-08-387-942C-4 from: 1 to: 997

```

Tue Apr 2 09:40:12 2002

```
11 ACTGATGATCTGAATTTGCTGCTGACGGATGGGGAAGACAACTAT 60
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
398 SerAspAlaHisGluThrLeuLeuGlyLeuAspGlyAsnAspArgLeuAs 414
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
61 AAGTGGTCTTACGAGGCTCAACAAGTNGTCCATCATCCACACAG 110
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
414 nGlyGlyAlaGlyAsnAspGlyAlaGlyAlaGlyArgAspAsnL 431
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
111 TCCTTTGGGCGCTCTGACGCTCAAGAACTAGAGGAGCTCTCCAAATG 160
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
431 euThrGlyAlaGlyAla.....AspLeuPheArgVal 442
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
161 ACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGAAATGGCT 210
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
443 SerAlaArgThrAspSerTyrArgThrAspSerAlaSerPheAsnAspLe 459
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
211 CATGATGCTTTGGGCGCTTTCATCAGGAATGAGGAGTGTCTCTCAGC 260
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
459 uileThrAspPheAspAla.....SerGlnA 468
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
261 GCTCCATCCAGCTTGAGTAAGGA 286
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
468 spArgileAspLeuSerAlaLeuGly 476
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
```

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:us-08-772-270A-11

```
seq_documentation_block:
; Sequence 11, Application US/08772270A
; Patent No. 6019984
; GENERAL INFORMATION:
; APPLICANT: MacInnes, Janet
; APPLICANT: Ricciatti, Paul
; APPLICANT: Mallard, Bonnie
; APPLICANT: Rosendal, Soren
; TITLE OF INVENTION: NOVEL BACTERIAL PREPARATIONS, METHOD FOR
; TITLE OF INVENTION: PRODUCING SAME, AND THEIR USE AS VACCINES
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bereskin & Parr
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/772,270A
; FILING DATE: December 23, 1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 6580-81
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1049 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Actinobacillus pleuropneumoniae
US-08-772-270A-11
```

alignment_scores:

```
Quality: 62.50 Length: 67
Ratio: 1.420 Gaps: 2
Percent Similarity: 65.672 Percent Identity: 29.851

alignment_block:
US-09-049-696-7 x US-08-772-270A-11 ..
Align_seg 1/1 to: US-08-772-270A-11 from: 1 to: 1049
44 GGSAGAGACAACTATAGTGGTGTCTTACGAGGTCAACAAAGTNG 93
   |||::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
981 GlyLysIleIleThrAsnAlaGlyAsnPheGlyValAlaLysGlnGlyAs 997
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
94 TGCCATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAACAACACTAG 143
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
997 nThrGlyIleAsnThrAlaAlaLeuAsn.....A 1007
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
144 AGGAGCTGTCCAAATCAGAGGAGGTTTACAGACATGCTTCAGATCAA 193
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1007 snGluValAsnIleIleSerSerAlaAsnThrPheAlaThrSerGln 1023
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
194 GTTCAGAACATGCGCTCATGTGCTTTGGGGCCCTTTCATCAGGAAA 243
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1024 LeuGlyGlySerGlyMet.....GlyThrLeuProSerThrAs 1036
   ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
244 T 244
1036 n 1036

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US93-10500-2

seq_documentation_block:
; Sequence 2, Application PC/TUS9310500
; GENERAL INFORMATION:
; APPLICANT: Chang, Yung-Fu
; TITLE OF INVENTION: Recombinant Vaccine For Procine
; TITLE OF INVENTION: Pleuropneumoniae
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Alan S. Korman
; STREET: 1600 Empire Tower
; CITY: Buffalo
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10500
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/972,229
; FILING DATE: 05-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Korman, Alan S.
; REGISTRATION NUMBER: 33,932
; REFERENCE/DOCKET NUMBER: 19603/00001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-853-8104
; TELEFAX: 716-853-8109
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1244 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
```

[illegible]

us-09-049-696-7.rai

Tue Apr 2 09:40:12 2002

This Page Blank (uspto)

OM of: US-09-049-696-7 to: A_Geneseq_1101.* out_format : pfs
 Date: Mar 30, 2002 2:18 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+2p.model -DEV=xlp
 -Q=/cgn2.1/USPTO_spool/US09049696/runat_28032002_145237_2001/app_query.fasta_1.12579
 -DB=A_Geneseq_1101 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
 -CAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=200000000
 -USER=US09049696 @CGN1_1_293 -NCPU=6 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:

Query: US-09-049-696-7
 Query length: 289
 Database: A_Geneseq_1101.*
 Database sequences: 522463
 Database length: 74073290
 Search time (sec): 819.910000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAV66749		315.00	694.62	1.1e-31	919
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAAB87560		315.00	694.62	1.1e-31	919
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAAB65272		315.00	694.62	1.1e-31	919
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAAB53310		304.00	684.06	1.8e-30	218
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAAG73761		304.00	684.06	1.8e-30	218
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAAY41039		218.00	474.96	3.0e-19	592
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAAB11324		218.00	474.96	3.0e-19	592
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAAB45904		218.00	472.62	3.2e-19	742
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAAY41040		218.00	471.96	3.3e-19	791
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAAB11325		218.00	471.96	3.3e-19	791
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAAY70459		218.00	470.16	3.5e-19	942
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAAY41036		218.00	470.15	3.5e-19	943
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAAB11321		218.00	470.15	3.5e-19	943
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAAB74823		213.00	458.59	1.5e-18	943
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAAB74827		117.00	274.21	1.1e-06	25
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAAG0884		80.50	156.43	0.1532	634
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAAB79158		80.50	156.43	0.1532	634
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAAB74302		70.50	131.22	3.17	777
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAAB43868		70.50	131.04	3.19	790
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAAB06520		70.50	130.46	3.25	836
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAAB76698		68.50	138.73	3.92	240
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAAG31604		68.00	133.74	5.13	348
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAAY37553		68.00	130.83	5.62	461
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAAB58241		67.50	138.62	4.92	194
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAAB41868		66.50	121.94	10.40	779
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAAY3596		66.50	121.94	10.40	779
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAAB41869		66.50	121.21	10.64	836
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAAB74676		65.00	129.74	11.37	262
/SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAAB80597		65.00	112.15	19.80	1437
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAAB10225		65.00	112.15	19.80	1437
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAAB47031		65.00	112.03	19.87	1453
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAAG91311		64.50	126.48	14.09	321
/SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT:AAAB54630		64.50	114.77	20.38	997
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAAY94309		64.00	120.62	18.96	506
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAAY43542		64.00	109.84	26.63	1437
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAAY38754		63.50	124.04	19.03	325

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAV66749

seq_documentation_block:

ID	AAV66749 standard; protein; 919 AA.
XX	AAV66749;
AC	AAV66749;
XX	AAV66749;
DT	05-APR-2000 (first entry)
XX	AAV66749;
DE	Membrane-bound protein PRO1124.
XX	Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW	pharmaceutical; receptor immunoadhesin; gene mapping.
OS	Homo sapiens
XX	AAV66749
PN	AAV66749
XX	AAV66749
PD	09-DEC-1999
XX	AAV66749
XX	AAV66749
PF	02-JUN-1999; 99WO-US12252.
XX	AAV66749
PR	02-JUN-1998; 98US-0087607.
PR	02-JUN-1998; 98US-0087609.
PR	02-JUN-1998; 98US-0087759.
PR	03-JUN-1998; 98US-0087827.
PR	04-JUN-1998; 98US-0088021.
PR	04-JUN-1998; 98US-0088025.
PR	04-JUN-1998; 98US-0088028.
PR	04-JUN-1998; 98US-0088029.
PR	04-JUN-1998; 98US-0088030.
PR	04-JUN-1998; 98US-0088033.
PR	04-JUN-1998; 98US-0088326.
PR	05-JUN-1998; 98US-0088167.
PR	05-JUN-1998; 98US-0088202.
PR	05-JUN-1998; 98US-0088212.
PR	05-JUN-1998; 98US-0088217.
PR	09-JUN-1998; 98US-0088655.
PR	10-JUN-1998; 98US-0088722.
PR	10-JUN-1998; 98US-0088730.
PR	10-JUN-1998; 98US-0088734.
PR	10-JUN-1998; 98US-0088738.
PR	10-JUN-1998; 98US-0088740.
PR	10-JUN-1998; 98US-0088741.
PR	10-JUN-1998; 98US-0088742.
PR	10-JUN-1998; 98US-0088810.
PR	10-JUN-1998; 98US-0088811.
PR	10-JUN-1998; 98US-0088824.
PR	10-JUN-1998; 98US-0088825.
PR	10-JUN-1998; 98US-0088826.
PR	10-JUN-1998; 98US-0088858.
PR	11-JUN-1998; 98US-0088861.
PR	11-JUN-1998; 98US-0088863.
PR	11-JUN-1998; 98US-0088866.
PR	12-JUN-1998; 98US-0088909.
PR	12-JUN-1998; 98US-0089105.
PR	16-JUN-1998; 98US-0089440.
PR	16-JUN-1998; 98US-0089512.
PR	16-JUN-1998; 98US-0089514.
PR	17-JUN-1998; 98US-0089532.
PR	17-JUN-1998; 98US-0089538.
PR	17-JUN-1998; 98US-0089598.
PR	17-JUN-1998; 98US-0089599.
PR	17-JUN-1998; 98US-0089600.
PR	17-JUN-1998; 98US-0089653.
PR	18-JUN-1998; 98US-0089801.
PR	18-JUN-1998; 98US-0089907.
PR	18-JUN-1998; 98US-0089908.
PR	19-JUN-1998; 98US-0089947.
PR	19-JUN-1998; 98US-0089948.
PR	19-JUN-1998; 98US-0089952.
PR	22-JUN-1998; 98US-0090246.
PR	22-JUN-1998; 98US-0090252.
PR	22-JUN-1998; 98US-0090254.
PR	23-JUN-1998; 98US-0090349.
PR	23-JUN-1998; 98US-0090355.

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.AAB87560

seq_documentation_block:

ID AAB87560 standard; Protein; 919 AA.

AC AAB87560;

XX 15-MAY-2001 (first entry)

XX Human PRO1124.

XX Human; PRO protein; mapping.

XX Homo sapiens.

XX WO200116318-A2.

XX 08-MAR-2001.

XX 24-AUG-2000; 2000WO-US23328.

XX 01-SEP-1999; 99WO-US20111.

XX 15-SEP-1999; 99WO-US21090.

XX 07-DEC-1999; 99US-0169495.

XX 09-DEC-1999; 99US-0170262.

XX 11-JAN-2000; 2000US-0175481.

XX 18-FEB-2000; 2000WO-US04341.

XX 18-FEB-2000; 2000WO-US04342.

XX 22-FEB-2000; 2000WO-US04414.

XX 01-MAR-2000; 2000WO-US05601.

XX 03-MAR-2000; 2000US-0187202.

XX 25-APR-2000; 2000US-0199397.

XX 22-MAY-2000; 2000WO-US14042.

XX 05-JUN-2000; 2000US-0209832.

XX (GETH) GENENTECH INC.

XX Eaton DL, Filvaroff E, Gerritsen ME, Coddard A, Godowski PJ;

PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;

XX WPI; 2001-183260/18.

XX N-PSDB; AAF92092.

XX Eighty four nucleic acids encoding PRO polypeptides, useful in

PT molecular biology, including use as hybridization probes, and in

PT chromosome and gene mapping.

XX Claim 12; Fig 70; 278pp; English.

XX The present sequence is a human PRO polypeptide (secreted and

CC transmembrane). The PRO protein, and PRO agonists, PRO antagonists or

CC anti-PRO antibodies are useful for preparation of a medicament useful in

CC the treatment of a condition which is responsive to the PRO protein,

CC agonists, antagonists or anti-PRO antibodies. The PRO protein may also be

CC employed as molecular weight markers for protein electrophoresis. The PRO

CC coding sequence has applications in molecular biology, including use as

CC hybridisation probes, and in chromosome and gene mapping.

XX Sequence 919 AA;

SQ

alignment_scores:

Quality: 315.00 Length: 92

Ratio: 3.889 Gaps: 0

Percent Similarity: 88.043 Percent Identity: 66.304

alignment_block:

US-09-049-696-7 x AAB87560 ..

Align seg 1/1 to: AAB87560 from: 1 to: 919

14 GATGGATCTGAATGTGCTCTGACGGATGGGAGACACACTATAAG 63

|||||
404 AspGlySerGluValLeuLeuThrAspGlyGluAspAsnThrAlaLase 420
-64 TGGGTGCTTTAAACAGAGTCAAAAGTNGTCCCATCATCCACACAGTCG 113
|||||
420 rSerCysIleAspGluValLysGlnSerGlyAlaIleValHisPheIleA 437
114 CTTTGGGCGCCTCTGACGCTCAAGAACTAGAGGAGCTGTCCAAATGACA 163
|||||
437 laLeuGlyArgAlaAlaAspGluAlaValIleGluMetSerLysIleThr 453
164 GGAGSTTTACAGACATATGCTTCAGATCAAGTTCAGACAAATGGCCTCAT 213
|||||
454 GlyGlySerHisPheTyrValSerAspGluAlaGlnAsnAsnGlyLeuIl 470
214 TGATGCTTTTGGGCGCCTTTTCATCAGGAAATGAGAGCTGTCTCTCAGCGCT 263
|||||
470 eAspAlaPheGlyAlaLeuThrSerGlyAsnThrAspLeuSerGlnLysS 487
264 CCATCCAGCTTGAGAGTAAGGATTAA 289
|||||
487 erLeuGlnLeuGluSerLysGlyLeu 495

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.AAB65272

seq_documentation_block:

ID AAB65272 standard; Protein; 919 AA.

AC AAB65272;

XX 02-APR-2001 (first entry)

XX Human PRO1124 (UNQ562) protein sequence SEQ ID NO:379.

XX Human; secreted and transmembrane protein; PRO; cytostatic;

XX cell death; cancer; chromosomal mapping; gene mapping; tissue typing;

XX diagnostic assay.

XX Homo sapiens.

XX WO200073454-A1.

XX 07-DEC-2000.

XX 30-MAR-2000; 2000WO-US08439.

XX 02-JUN-1999; 99WO-US12252.

XX 23-JUN-1999; 99US-0141037.

XX 07-JUL-1999; 99US-0143048.

XX 20-JUL-1999; 99US-0144758.

XX 26-JUL-1999; 99US-0145698.

XX 28-JUL-1999; 99US-0146222.

XX 17-AUG-1999; 99US-0149396.

XX 15-SEP-1999; 99WO-US21090.

XX 15-SEP-1999; 99WO-US21547.

XX 08-OCT-1999; 99US-0158663.

XX 30-NOV-1999; 99WO-US28313.

XX 01-DEC-1999; 99WO-US30095.

XX 16-DEC-1999; 99WO-US30911.

XX 05-JAN-2000; 2000WO-US00219.

XX 06-JAN-2000; 2000WO-US00376.

XX 11-FEB-2000; 2000WO-US03565.

XX 18-FEB-2000; 2000WO-US04341.

XX 22-FEB-2000; 2000WO-US04414.

XX 24-FEB-2000; 2000WO-US04914.

XX 24-FEB-2000; 2000WO-US05004.

XX 02-MAR-2000; 2000WO-US05841.

XX 15-MAR-2000; 2000WO-US06884.

XX 20-MAR-2000; 2000WO-US07377.

XX (GETH) GENENTECH INC.

XX

PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrarini N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Kijavini IJ, Napier MA, Pan J, Paoni NF;
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
PI Zhang Z;
DR WPI: 2001-032160/04.
DR N-PSDB: AAF44241.
XX
XX PRO polynucleotides used to produce polypeptides used to target
PT bioactive molecules such as toxins, radiolabels or antibodies, to
PT specific cells, to cause targeted cell death -
XX
XX Claim 12; Fig 274; 935pp; English.
XX
XX The present invention describes human secreted and transmembrane PRO
CC proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC can be used for targeted delivery of bioactive molecules, such as
CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC sequences, and their fragments, can be used as hybridisation probes, in
CC chromosomal and gene mapping, and in the generation of anti-sense RNA
CC and DNA. They may also be used to produce transgenic animals which are
CC used to develop and screen therapeutically useful reagents. The PRO
CC nucleotide and protein sequence can be used for tissue typing and in
CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
CC in the isolation of human PRO sequences. AAF44087 to AAF44269 and
CC AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC sequences given in the exemplification of the present invention.
XX
SQ Sequence 919 AA;

alignment_scores:
Quality: 315.00 Length: 92
Ratio: 3.889 Gaps: 0
Percent Similarity: 88.043 Percent Identity: 66.304
alignment_block:
US-09-049-696-7 x AAB65272 ..
Align seg 1/1 to: AAB65272 from: 1 to: 919
14 GATGGATCTGAATTTGTGCTGACGGATGGGAGAGACAACACTATAAG 63
||||| :
404 AspGlySerGluValLeuLeuLeuLeuLeuThrAspGlyGluAspAsnThrAlaSe 420
64 TGGGTGCTTTAACGAGTCAAAAGTGTGCTATCCACACAGTCG 113
| :
420 rSerCysIleAspGluValLeuGlnSerGlyAlaIleValHisPheIleA 437
114 CTTTGGGGCCCTCTGCAGCTCAAGACTAGAGGAGCTGTCCAAATGACA 163
||||| :
437 laLeuGlyArgAlaAlaAspGluAlaValIleGluMetSerLysIleThr 453
164 GGAGGTTTACAGACATATGCTTCAGATCAAGTTTCAGACAAATGCCCTCAT 213
||||| :
454 GlyGlySerHisPheTyrlValSerAspGluAlaGlnAsnGlyLeuII 470
214 TGATGCTTTGGGGCCCTTCATCAGGAAATGGAGCTGTCTCTCAGCGCT 263
||||| :
470 eaSpAlaPheGlyAlaLeuThrSerGlyAsnThrAspLeuSerGlnLys 487
264 CCATCCAGCTTGAGAGTAAGGATTA 289
| :
487 erLeuGlnLeuGluSerLysGlyLeu 495

seq_name: /SDS2/gcgdata/geneseq/geneseq/AA2000.DAT: AAB53310

seq_documentation_block:
ID AAB53310 standard; Protein; 218 AA.
XX
AC AAB53310;

XX
DT 09-MAR-2001 (first entry)
DE Human colon cancer antigen protein sequence SEQ ID NO:850.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.
XX
OS Homo sapiens.
XX
XX WC2000055351-A1.
PN
XX 21-SEP-2000.
PD
XX
XX 08-MAR-2000; 2000WO-US05883.
PF
XX
XX 12-MAR-1999; 99US-0124270.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM;
PI
XX
XX WPI: 2000-587534/55.
DR
XX
XX N-PSDB: AAC98067.
XX
XX
XX Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
XX
XX Claim 11; Page 1400; 2104pp; English.
XX
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB533234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal, and
CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 218 AA;

alignment_scores:
Quality: 304.00 Length: 92
Ratio: 3.848 Gaps: 0
Percent Similarity: 85.870 Percent Identity: 64.130
alignment_block:
US-09-049-696-7 x AAB53310 ..

Align seg 1/1 to: AAB53310 from: 1 to: 218

14 GATGGATCTGAATTTGTGCTGACGGATGGGAGAGACAACACTATAAG 63
||||| :
24 AspGlySerGluValLeuLeuLeuLeuThrAspGlyGluAspAsnThrAlaSe 40
64 TGGGTGCTTTAACGAGTCAAAAGTGTGCTATCCACACAGTCG 113
| :
40 rSerCysIleAspGluValLeuGlnSerGlyAlaIleValHisPheIleA 57

[illegible]

The invention provides isolated human lung tumor nucleic acids and polypeptides. The polypeptides can be used for the treatment of lung cancer. The polypeptides and polynucleotides can be used to stimulate T cells or antigen presenting cells for use in the treatment of lung cancer. The polypeptides and monoclonal antibodies specific for the polypeptides can also be used to inhibit the development of lung cancer. Agents which bind the polypeptides can be used for detecting lung cancer, and for monitoring the progression of lung cancer.

```

SU sequence 791 AA:
      *
      alignment_scores:
      ..      Quality: 218.00      Length: 90
      ..      Ratio: 3.028      Gaps: 0
      ..      Percent Similarity: 80.000      Percent Identity: 47.778

      alignment_block:
      ..      US-09-049-696-7 x ANY41040      ..
      ..      Align seq 1/1 to: ANY41040      from: 1 to: 791

```

[illegible]

...
CC are treated with P2, polynucleotides encoding P2 or antigen presenting
CC cells expressing P2 and then administered to the patient to inhibit
CC development of cancer.

FT Modified-site 368 /note= "Phosphorylation site"
 FT Modified-site 418 /note= "Phosphorylation site"
 FT Modified-site 419 /note= "Phosphorylation site"
 FT Modified-site 474 /note= "Phosphorylation site"
 FT Modified-site 498 /note= "Phosphorylation site"
 FT Modified-site 797 /note= "Phosphorylation site"
 FT Modified-site 2 /note= "Phosphorylation site"
 FT Modified-site 76 /note= "Phosphorylation site"
 FT Modified-site 92 /note= "Phosphorylation site"
 FT Modified-site 270 /note= "Phosphorylation site"
 FT Modified-site 318 /note= "Phosphorylation site"
 FT Modified-site 384 /note= "Phosphorylation site"
 FT Modified-site 557 /note= "Phosphorylation site"
 FT Modified-site 568 /note= "Phosphorylation site"
 FT Modified-site 593 /note= "Phosphorylation site"
 FT Modified-site 652 /note= "Phosphorylation site"
 FT Modified-site 685 /note= "Phosphorylation site"
 FT Modified-site 822 /note= "Phosphorylation site"
 FT Modified-site 823 /note= "Phosphorylation site"
 FT Modified-site 838 /note= "Phosphorylation site"
 FT Modified-site 927 /note= "Phosphorylation site"
 FT Modified-site 74 /note= "Phosphorylation site"
 FT Modified-site 97 /note= "Glycosylation site"
 FT Modified-site 150 /note= "Glycosylation site"
 FT Modified-site 231 /note= "Glycosylation site"
 FT Modified-site 235 /note= "Glycosylation site"
 FT Modified-site 253 /note= "Glycosylation site"
 FT Modified-site 291 /note= "Glycosylation site"
 FT Modified-site 521 /note= "Glycosylation site"
 FT Modified-site 555 /note= "Glycosylation site"
 FT Modified-site 579 /note= "Glycosylation site"
 FT Modified-site 636 /note= "Glycosylation site"
 FT Modified-site 821 /note= "Glycosylation site"
 FT Modified-site 937 /note= "Glycosylation site"
 FT Peptide 1..31 /label= Signal_peptide
 FT Domain 900..926 /label= Transmembrane_domain
 FT XX

PN WO200012711-A2.
 XX 09-MAR-2000..
 XX 02-SEP-1999; 99WO-US20468.
 XX 02-SEP-1998; 98US-0145815.
 PR 12-NOV-1998; 98US-0191283.
 PR 09-DEC-1998; 98US-0208821.
 PR 26-JAN-1999; 99US-0237506.
 PR 10-FEB-1999; 99US-0247891.
 XX (INCY-) INCYTE PHARM INC.
 PA AuYoung J, Bandman O, Tang YT, Reddy R, Hillman JL, Yue H;
 PI Lal P, Corley NC, Guegler KJ, Gorgone G, Baughn MR, Azimzai Y;
 XX WPI: 2000-256643/22.
 DR N-PSDB; AA251625.
 XX
 PT Novel human membrane channel protein and polynucleotide useful for
 PT diagnosing and treating cell proliferative, inflammatory, secretory,
 PT osmoregulatory, muscular, cardiovascular and neurological disorders
 XX
 PS Claim 1; Page 107-110; 140pp; English.
 CC The present sequence is the human membrane channel protein-9 (MECHP-9),
 CC which is expressed in dermatologic, gastrointestinal and reproductive
 CC tissues. Anti-MECHP antibodies can be used as therapeutic antagonists and
 CC reagents for diagnosis and monitoring diseases. MECHP cDNA can be used
 CC for diagnosis of MECHP-related diseases and gene mapping. MECHP can be
 CC used for treatment of cell proliferative disorders such as bursitis and
 CC atherosclerosis, cancers like lymphoma and sarcoma, inflammatory
 CC disorders like AIDS and Addison's disease, transport/secretory disorders
 CC like cystic fibrosis and diabetes mellitus, osmoregulatory disorders like
 CC diarrhoea and renal failure, muscular disorders like myocarditis and
 CC Duchenne's muscular dystrophy, cardiovascular disorders like hypertension
 CC and vasculitis, congenital lung anomalies like bronchitis and asthma and
 CC neurological disorders like Alzheimer's disease, Parkinson's disease and
 CC Huntington's disease.
 XX
 SQ Sequence 942 AA;

 alignment_scores:
 Quality: 218.00 Length: 90
 Ratio: 3.028 Gaps: 0
 Percent Similarity: 80.000 Percent Identity: 47.778

 alignment_block:
 US-09-049-696-7 x AAY70459 ..
 Align seg 1/1 to: AAY70459 from: 1 to: 942
 17 GGATCTCAAAATTGCTGCTGACGGATGGGAGACACACTATAAGTGG 66
 411 GlySerValMetIleLeuValThrSerGlyAspLysLeuLeuGlyAs 427
 67 GTGCTTTAACGAGGTCAACAAAGTNGTCATCATCCACACAGTCGCTT 116
 427 nCysLeuProThrValLeuSerSerGlySerThrIleHisSerIleAlaL 444
 117 TGGGGCCCTCTGCAGCTCAAGACTAGAGGAGCTGTCCAAAATGACAGGA 166
 444 euGlySerSerAlaAlaProAsnLeuGluLeuSerArgLeuThrGly 460
 167 GGTTCACAGACATATGCTTCAGATCAAGTTCAGAACAAATGCCCTCATGA 216
 461 GlyLeuLysPhePheValProAspIleSerAsnSerAsnSerMetIleAs 477
 217 TGCTTTTGGGGCCCTTTCATCAGGAATGGAGCTGCTCTCAGCGCTCCA 266
 477 pAlaPheSerArgIleSerGlyThrGlyAspIlePheGlnGlnHisI 494


```
alignment_scores:      Quality: 218..00    Length: 90
                       Ratio: 3.028       Gaps: 0
Percent Similarity: 80.000   Percent Identity: 47.778

alignment_block:
US-09-045-696-7 x AAB11321 ..

Align seg 1/I to: AAB11321 from: 1 to: 943

17 GGATTTGAAATGTGCTGCACGGATGGGGAAGACAACACTATAAGTGG 66
||||| :|||:|||||:|||||:|||||:|||||:|||||:
412 GlySerValMetIleLeuValThrSerGlyAspAspLysLeuLeuGlyAs 428

67 GTGCTTTAACGAGGTCAAACAAGTNGTGCCATCATCCACACAGTCGCWT 116
|||||: ||| :|||: ||| :|||: |||: |||: |||: |||: |||:
428 nCysLeuProThrValLeuSerSerGlySerThrIleHisSerIleAlaL 445

117 TGGGGCCCTCTGCAGCTCAAGAAGTAGAGGAGCTGCCAAAATGCACAGA 166
||||| ||||| ||||| :|||: ||||| |||||: |||||: |||||:
445 euGlySerSerAlaAlaProAsnLeuGluLeuSerArgLeuThrGly 461

167 GGTTTTACAGACATATGCTTCAGATCATAGTTWCAACAATGGCGTCAATGA 216
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
462 GlyLeuLysPhePheValProAspIleSerAsnSerAsnSerMetIleAs 478

217 TGCTTTTGGGGCCCTTTTCATCAGGAATGGAGCTGCTCTCAGCGGCTCCA 266
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
478 palApeSerArgIleSerSerGlyThrGlyAspIlePheGlnGlnHisI 495

267 TCCAGCTTGAGAGTAAGGGA 286
||||| ||||| ||||| |||
495 LeGlnLeuGluSerThrGly 501
```

seq_name: /SIDS2/gcgcdata/geneseq/geneseq/AA1999.DAT: AAB74823

seq_documentation_block:	
ID	AAB74823 standard; Protein: 943 AA.
XX	
AC	AAB74823;
XX	
DT	13-JUN-2001 (first entry)
XX	
DE	Human ICACC-2 protein sequence.
XX	
KW	ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW	interleukin 9 induced calcium activated chloride channel; IL-9;
KW	calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW	anti-inflammatory; immunomodulatory; cystic fibrosis;
KW	inflammatory bowel disease; autoimmune disease.
XX	
OS	Homo sapiens.
XX	
PN	W09944620-A1.
XX	
PD	10-SEP-1999.
XX	
PF	03-MAR-1999; 99WO-US04703.
XX	
PR	03-MAR-1998; 98US-0076815.
XX	
PA	(MAGA-) MAGAININ PHARM INC.
XX	
PI	Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;
PI	Nicolaides NC, Zhou Y, Dong Q;
XX	
WPI	1999-550979/46.
DR	N-PSDB; AAF81926.
XX	
PT	New nucleic acid encoding calcium activated chloride channel, used to
PT	identify, e.g. specific modulators for treating atopic allergy -
XX	
PS	Claim 13; Fig 4A; 75pp; English.

The present sequence represents the human interleukin 9 (IL-9) induced calcium activated chloride channel 2 (ICACC-2) protein. ICACC proteins have anti-allergic, anti-asthmatic, anti-inflammatory and immunomodulatory activities. Compounds (A) that downregulate ICACC used to alleviate asthma (or more generally atopic allergy), while those (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and inflammatory bowel disease (IBD) (or other autoimmune diseases). Measuring levels of ICACC-1 can be used in the diagnosis of asthma (increased levels) or IBD (reduced levels), also for monitoring treatment of these conditions. The ICACC proteins can be used: (A) to raise specific antibodies (Ab), useful: (a) as immunoassay reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as (A) to disrupt bonding between ICACC and its ligands; and (iii) to identify modulators and binding partners. ICACC polynucleotides can be used to generate transgenic animals or recombinant cells, used to screen for antagonists, also as a source of therapeutic antisense agents or diagnostic probes (for quantifying mRNA expression, e.g. for identification of modulators).

Sequence 943 AA:

Alignment scores:
Quality: 213.00 Length: 90
Ratio: 2.958 Gaps: 0
Percent Similarity: 80.000 Percent Identity: 46.667

alignment_block:
US-09-049-696-7 x AAB74823 ..

Align seg 1/1 to: AAB74823 from: 1 to: 943

17 GGATCTGAATAATTGCTGCAGCGATGGGGAACAACACTATAAGTGG 66
||||| :||::|||::|||::|||::|||::|||::|||::|||::|||::
412 GlySerValMetIleLeuValThrSerGlyAspAspLysLeuGlyAs 428

67 GTGCTTTAACGAGGTCAAAAGAAGTNGTGCCATCATCCACACAGTCGCTT 116
::|||:: ||| ::|||:: ||| ::|||::|||::|||::|||::|||::
429 nCysLeuProthrValLeuSerSerGlySerThrIleHisSerIleAlaL 445

117 TGGGGCCCCCTGCAGCTCAGAACTAGACGAGCTGTCCAANTGACAGGA 166
|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
445 euGlySerSerAlaAlaProAsnLeuHisGluLeuSerArgLeuThrGly 461

167 GGTTCACAGACATATGCTTCAGATCAAGTTCAGAAACATGGCTCATTTGA 216
|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
462 GlyLeuLysPhePheValProAspIleSerAsnSerAsnSerMetIleAs 478

217 TGCTTTTGGGGCCCTTTTCATCAGAAATGGAGCTGTCTCTCACGCCCTCCA 266
|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
478 pAlaPheSerArgIleSerSerGlyThrGlyAspIlePheGlnGlnHisI 495

267 TCCAGCTTGAGATGAAGGGA 286
|||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
495 leGlnLeuGlusSerThrGly 501

seq_name: /SDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:AAB74827

seq_documentation_block:
ID ID AAB74827 standard; Peptide; 25 AA.
XX XX
XX AAB74827;
XX XX
DT DT 13 JUN-2001 (first entry)
DE DE
DE DE
KW KW ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW KW interleukin 9 induced calcium activated chloride channel; IL-9;
KW KW calcium activated chloride channel; anti-allergic; anti-asthma;
KW KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW KW inflammatory bowel disease; autoimmune disease

XX Mus sp.
OS WO9944620-A1.
XX 10-SEP-1999.
XX 03-MAR-1999; 99WO-US04703.
XX 03-MAR-1998; 98US-0076815.
XX (MAGA-) MAGAININ PHARM INC.
XX Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;
PI Nicolaides NC, Zhou Y, Dong Q;
XX WPI: 1999-550979/46.
XX
XX New nucleic acid encoding calcium activated chloride channel, used to
PT identify, e.g. specific modulators for treating atopic allergy -
XX
XX Claim 24; Page 41; 75pp; English.
XX
XX The present invention describes interleukin 9 (IL-9) induced
CC calcium activated chloride channel (ICACC) proteins. ICACC proteins
CC have anti-allergic, anti-asthmatic, anti-inflammatory and
CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
CC used to alleviate asthma (or more generally atopic allergy), while those
CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
CC (increased levels) or IBD (reduced levels), also for monitoring
CC treatment of these conditions. The ICACC proteins can be used:
CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
CC reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
CC identify modulators and binding partners. ICACC polynucleotides can be
CC used to generate transgenic animals or recombinant cells, used to screen
CC for antagonists, also as a source of therapeutic antisense agents or
CC diagnostic probes (for quantifying mRNA expression, e.g. for
CC identification of modulators). The present sequence represents a peptide
CC fragment of the murine ICACC-1 protein, which is used in the
CC exemplification of the present invention.
XX
XX Sequence 25 AA;

alignment_scores: Quality: 117.00 Length: 23
 Ratio: 5.087 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 95.652

alignment_block:
US-09-049-696-7 x AAB74827 ..

Align seg 1/1 to: AAB74827 from: 1 to: 25

2 AAATATCCACTGATGATCTGAATGTGCTGCTGACGGATGGGAGA 51
|||||
3 LysTyrProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAs 19
|||||
52 CAACACTATAAGTGGTGCC 70
|||||
19 pAsnThrIleSerSerCys 25

OM of: US-09-049-696-6 to: SPTREMBL_17.* out_format : pfs
 Date: Mar 30, 2002 2:46 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
 -Q/cgn2.1/USPTO_spool/US09049696/runat_28032002_145238_2053/app_query.fasta_1.12579
 -DB=SPTREMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -GAPOP=4.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DEPOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=dlsumg2
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=200000000
 -USER=US09049696_ecgn1_1_705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:

Query: US-09-049-696-6
 Query length: 252
 Database: SPTREMBL_17.*
 Database sequences: 473505
 Database length: 146272329
 Search time (sec): 805.760000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
sp_human:O95151	+	418.00	883.20	7.0e-42	914	O95151 homo sapiens (human). ca
sp_human:O9UPC6	+	418.00	883.20	7.0e-42	914	O9UPC6 homo sapiens (human). ca
sp_human:O9UNF6	+	418.00	883.20	7.0e-42	914	O9UNF6 homo sapiens (human). ca
sp_mammal:O9TUB5	+	337.00	708.02	4.0e-32	917	O9TUB5 sus scrofa (pig). epithe
sp_rodent:O88826	+	330.00	692.91	2.8e-31	913	O88826 mus musculus (mouse). gc
sp_rodent:O9D726	+	330.00	692.91	2.8e-31	913	O9D726 mus musculus (mouse). ch
sp_mammal:O9UNF7	+	268.00	558.81	8.2e-24	917	O9UNF7 homo sapiens (human). ca
sp_mammal:O29282	+	258.50	552.76	1.5e-22	109	O29282 sus scrofa (pig). unknow
sp_rodent:O88860	+	248.00	515.68	2.1e-21	902	O88860 mus musculus (mouse). ch
sp_rodent:O9QX15	+	248.00	515.68	2.1e-21	902	O9QX15 mus musculus (mouse). ca
sp_rodent:O9R070	+	245.00	509.19	4.8e-21	902	O9R070 mus musculus (mouse). ca
sp_rodent:O9EQR4	+	245.00	509.19	4.8e-21	902	O9EQR4 mus musculus (mouse). en
sp_mammal:O18742	+	233.00	484.11	1.4e-19	794	O18742 bos taurus (bovine). lu
sp_mammal:O18743	+	233.00	483.89	1.4e-19	820	O18743 bos taurus (bovine). lu
sp_mammal:O18741	+	233.00	483.22	1.3e-19	905	O18741 bos taurus (bovine). lu
sp_human:O9UQC9	+	186.00	381.30	6.1e-14	943	O9UQC9 homo sapiens (human). ch
sp_human:O9V6N2	+	186.00	381.30	6.1e-14	943	O9V6N2 homo sapiens (human). ca
sp_bacteria:O9RJ18	+	83.00	160.88	0.1630	672	O9RJ18 streptomyces coelicolor.
sp_invertebrate:O9VNX0	-	79.00	146.83	0.4467	1485	O9VNX0 drosophila melanogaster
sp_bacteria:O9JY15	+	74.00	145.74	2.14	356	O9JY15 neisseria meningitidis
sp_bacteria:O9JT16	+	74.00	145.74	2.14	356	O9JT16 neisseria meningitidis
sp_invertebrate:O9U7P4	+	73.50	142.43	2.36	494	O9U7P4 eufolliculina uhligi.
sp_bacteria:O9KMZ5	+	71.50	136.28	3.98	646	O9KMZ5 vibrio cholerae. hypothe
sp_rodent:O9CW45	-	71.00	142.40	5.23	224	O9CW45 mus musculus (mouse). 13
sp_rodent:O35802	+	70.50	131.62	5.01	932	O35802 rattus norvegicus (rat).
sp_organelle:O9G5W9	-	70.00	137.38	6.54	341	O9G5W9 trapezus savignyi. nath
sp_plant:O9ZV10	-	70.00	132.60	5.98	689	O9ZV10 arabidopsis thaliana (mc
sp_human:O9H7W0	-	69.00	137.69	9.04	237	O9H7W0 homo sapiens (human). cd
sp_mammal:O18744	-	69.00	135.20	8.63	342	O18744 bos taurus (bovine). lu
sp_fungi:O9V7P3	-	69.00	132.44	8.19	513	O9V7P3 schizosaccharomyces pont
sp_human:O9NZU0	-	69.00	130.84	7.95	649	O9NZU0 homo sapiens (human). ki
sp_human:O9P259	-	69.00	130.71	7.93	662	O9P259 homo sapiens (human). ki
sp_invertebrate:O9VT34	-	69.00	128.70	7.64	889	O9VT34 drosophila melanogaster
sp_bacteria:O05073	+	68.50	135.90	10.25	263	O05073 streptomyces clavuligeru
sp_invertebrate:O9V6W2	+	68.50	127.57	8.77	895	O9V6W2 drosophila melanogaster
sp_invertebrate:O04588	+	68.00	127.93	10.35	724	O04588 eimeria maxima. micron
sp_invertebrate:O18924	-	67.00	128.98	14.50	452	O18924 caenorhabditis elegans
sp_human:O13586	+	67.00	126.15	13.75	685	O13586 homo sapiens (human). gc
sp_archaea:O9Y9L4	-	67.00	123.58	13.10	999	O9Y9L4 aeropyrum pernix. 999aa
sp_bacteria:O33957	+	66.50	119.25	14.16	1611	O33957 streptomyces fradiae. d

sp_bacteria:O33956 + 66.50 113.54 12.72 3729 ! O33956 streptomyces fradiae
 sp_bacteria:O33954 + 66.50 112.30 12.43 4472 ! O33954 streptomyces fradiae
 sp_plant:O9SXD6 + 66.00 129.45 20.10 307 ! O9SXD6 arabidopsis thaliana
 sp_fungi:O9C229 - 65.00 117.92 18.98 1425 ! O9C229 neurospora crassa. r
 sp_bacteria:O9PKG2 - 65.00 133.50 29.81 123 ! O9PKG2 chlamydia muridarum.
 seq_name: sp_human:O95151

seq_documentation_block:
 ID O95151 PRELIMINARY; PRT; 914 AA.
 AC O95151;
 DT 01-MAY-1999 (TREMREL. 10, Created)
 DT 01-MAY-1999 (TREMREL. 10, Last sequence update)
 DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
 DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
 GN HCLCAL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SMALL INTESTINE;
 RX MEDLINE=99047526; PubMed=9828122;
 RA Gruber A.D., Elble R.C., Ji H.L., Schreuer K.D., Fuller C.M.,
 KA Pauli B.U.;
 RT "Genomic cloning, molecular characterization, and functional analysis
 of human HCLCAL, the first human member of the family of Ca2+-activated
 Cl-channel proteins.";
 RL Genomics 54:200-214(1998).
 DR EMBL; AF039400; AAC95428.1;
 DR InterPro; IPR000131; ATPase_gamma.
 DR InterPro; IPR002035; VWFA.
 DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
 DR PROSITE; PS50234; VWFA; 1.
 DR SMART; SM00327; VWA; 1.
 SQ SEQUENCE 914 AA; 100174 MW; 7066440C46557FA3 CRC64;

alignment_scores:
 Quality: 418.00 Length: 84
 Ratio: 4.976 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-049-696-6 x O95151
 Align seg 1/1 to: O95151 from: 1 to: 914

1 CAAGAATTGTTGTTTAGTCTTTCGACAAATCTGGAAGCATGGCGACTGG 50
 .. |||||.....
 304 GlnArgIleValCysLeuValLeuAspLysSerGlySerMetAlaThrGl 320
 .. |||||.....
 51 TAACCGCTCAATCGACTGAATCAAGCAGGCGAGCTTTCTCTGCTGCAG 100
 .. |||||.....
 320 YAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnPheLeuLeuGlnT 337
 .. |||||.....
 101 CAGTTGAGCTGGGTCTCGGTTCGGGTGCGAGTGTGACATTTGACAGTCTGCC 150
 .. |||||.....
 337 hrValGluLeuGlySerTrpValGlyMetValThrPheAspSerAlaAla 353
 .. |||||.....
 151 CATGTACAAAGTGAACATCATACAGATAAACAGTGGCAGTGACAGGACAC 200
 .. |||||.....
 354 HisValGlnSerGluLeuIleGlnIleAsnSerGlySerAspArgAspTh 370
 .. |||||.....
 201 ACTGCCAAAAGATTACTTCGACGAGCTTCAGGAGGAGGAGGAGGAGGAGG 250
 .. |||||.....
 370 rLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyThrSerIleCys 387
 .. |||||.....
 251 GC 252
 .. ||
 387 er 387

Tue Apr 2 09:40:11 2002

us-09-049-696-6.ispt

```

seq_name: sp_human:Q9UPC6

seq_documentation_block:
ID Q9UPC6 PRELIMINARY; PRT; 914 AA.
AC Q9UPC6;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCA1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Elbie R.C., Ji H.L., Schreier K.D., Fuller C.M.,
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
RT of human CLCA1, the first human member of the family of Ca2+-activated
RT Cl-channel proteins."
RL Genomics 54:200-214(1998).
DR EMBL: AF039401; AAC95429.1; -.
DR InterPro: IPR000131; ATPase_gamma.
DR PROSITE: PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE: PS0234; VWFA; 1.
DR SMART: SM00327; VWA; 1.
SQ SEQUENCE 914 AA; 100160 MW; 7761362B66752AA3 CRC64;

alignment_scores:
Quality: 418.00 Length: 84
Ratio: 4.976 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-6 x Q9UPC6 ..
Align seg 1/1 to: Q9UPC6 from: 1 to: 914

1 CAAGAATGTGTAGTTCCTTGACAAATCTGGAAGCATGGGACTGG 50
304 GlnArgIleValCysLeuValLeuAspLysSerGlySerMetAlaThrGI 320

51 TAACGGCTCAATCAGTGAATCAAGCAGGCGAGCTTTCTGCTGCAGA 100
320 YasnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheLeuGlnT 337

101 CAGTTGAGTGGGTCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGC 150
337 hrValGluLeuGlySerTrpValGlyMetValThrPheAspSerAlaAla 353

151 CATGTACAAGTCACTCATACAGATAAAGTGGCAGTGCAGGAGCAC 200
354 HisValGlnSerGluLeuIleGlnIleAsnSerGlySerAspArgAspTh 370

201 ACTCGCAAAAGATTACCTGCAGCAGCTTCAGAGGAGGAGCTCCATCTGCA 250
370 rLeuAlaLysArgLeuProAlaAlaSerGlyGlyThrSerIleCys 387

251 GC 252
387 er 387

seq_name: sp_mammal:Q9TUB5

seq_documentation_block:
ID Q9TUB5 PRELIMINARY; PRT; 917 AA.
AC Q9TUB5;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN.
GN AECC.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;

seq_name: sp_human:Q9UNF6

seq_documentation_block:
ID Q9UNF6 PRELIMINARY; PRT; 914 AA.
AC Q9UNF6;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

```

```

DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 1.
GN CACCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99364503; PubMed=10437792;
RA Agnel M., Vermet T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
RT chloride channel (CaCC) family predominantly expressed in the
RT digestive tract and trachea."
RL FEBS Lett. 455:295-301(1999).
DR EMBL: AF127036; AAD25487.1; -.
DR InterPro: IPR000131; ATPase_gamma.
DR PROSITE: PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE: PS0234; VWFA; 1.
DR PROSITE: PS0234; VWFA; 1.
SQ SEQUENCE 914 AA; 100202 MW; B58B451D97887D6B CRC64;

```

```

alignment_scores:
Quality: 418.00 Length: 84
Ratio: 4.976 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-6 x Q9UNF6 ..
Align seg 1/1 to: Q9UNF6 from: 1 to: 914

```

```

1.CAAAGAATGTGTAGTTCCTTGACAAATCTGGAAGCATGGGACTGG 50
304 GlnArgIleValCysLeuValLeuAspLysSerGlySerMetAlaThrGI 320

51 TAACGGCTCAATCAGTGAATCAAGCAGGCGAGCTTTCTGCTGCAGA 100
320 YasnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheLeuGlnT 337

101 CAGTTGAGTGGGTCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGC 150
337 hrValGluLeuGlySerTrpValGlyMetValThrPheAspSerAlaAla 353

151 CATGTACAAGTCACTCATACAGATAAAGTGGCAGTGCAGGAGCAC 200
354 HisValGlnSerGluLeuIleGlnIleAsnSerGlySerAspArgAspTh 370

201 ACTCGCAAAAGATTACCTGCAGCAGCTTCAGAGGAGGAGCTCCATCTGCA 250
370 rLeuAlaLysArgLeuProAlaAlaSerGlyGlyThrSerIleCys 387

251 GC 252
387 er 387

```

```
seq_name: sp_mammal:Q9TUB5
```

```

seq_documentation_block:
ID Q9TUB5 PRELIMINARY; PRT; 917 AA.
AC Q9TUB5;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN.
GN AECC.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;

```

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=ILEAL MUCOSA;
 RA Gaspar K.J., Gabriel S.E., Racette K.J., Forsyth G.W.;
 RT "A cDNA involved in porcine exocrine chloride conductance";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF095584; RA000077.1;
 DR InterPro: IPR000131; ATPhase_gamma.
 DR InterPro: IPR002035; VWFA.
 DR PROSITE: PS00153; ATPASE_GAMMA; UNKNOWN_1.
 DR PROSITE: PS0234; VWFA; 1.
 DR SMART: SM00327; VWFA; 1.
 SQ SEQUENCE 917 AA; 100735 MW; 649F41944F7EAD19 CRC64;

alignment_scores:
 US-09-049-696-6 x Q9TUB5 ..
 Align seg 1/1 to: Q9TUB5 from: 1 to: 917
 1 CAAGAATTGCTGTTAGTCTTGCACAAATCTGGAAGCATGGCGACTGG 50
 |||||
 304 GlnArgIleValCysLeuValLeuAspLysSerGlySerMetThrValGI 320
 51 TAACGGCTCAATCGACTGAATCAAGCAGGCGAGCTTTCTCTGCTGCAGA 100
 |||||
 320 yGlyArgLeuLysArgLeuAsnGlnAlaGlyLysLeuPheLeuLeuGlnT 337
 101 CAGTTGAGCTGGGCTCTGGGATGGTGATGACATTGACAGTGGCTGCC 150
 |||||
 337 hrValGluGlnGlyAlaTrpValGlyMetValAlaPheAspSerAlaAla 353
 151 CATGTACAAAGTGAATCATACAGATAAAGAGTGGCAGTGCAGGGACAC 200
 |||||
 354 TyrValLysSerGluLeuValGlnIleAsnSerAlaAlaGluArgAspAl 370
 201 ACTGCCAAAGATTACTCGCAGCAGCTTCAGGAGGAGGACGTCATCTGCA 250
 |||||
 370 aLeuAlaArgSerLeuProThrAlaAlaSerGlyGlyThrSerIleCysS 387
 251 GC 252
 ||
 387 er 387

seq_name: sp_rodent:O88826

seq_documentation_block:
 ID O88826 PRELIMINARY; PRT; 913 AA.
 AC O88826;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE GOB-5 PROTEIN.
 GN GOB-5 OR GOB-5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=INTESTINE;
 RA Komiyama T., Tanigawa Y., Hirohashi S.;
 RT "Cloning of a gene, gob-5, which is expressed in intestinal goblet cells in mice."
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB017156; BAA33743.1;
 DR MGI: 1346342; Clca3
 DR InterPro: IPR000131; ATPase_gamma.

DR InterPro: IPR002035; VWFA.
 DR SMART: SM00327; VWFA; 1.
 DR PROSITE: PS00153; ATPASE_GAMMA; UNKNOWN_1.
 DR PROSITE: PS0234; VWFA; 1.
 SQ SEQUENCE 913 AA; 100070 MW; A7FA2F9E1089806D CRC64;
 alignment_scores:
 US-09-049-696-6 x O88826 ..
 Align seg 1/1 to: O88826 from: 1 to: 913
 1 CAAGAATTGCTGTTAGTCTTGCACAAATCTGGAAGCATGGCGACTGG 50
 |||||
 305 GlnArgIleValCysLeuValLeuAspLysSerGlySerMetLeuAsnAs 321
 51 TAACGGCTCAATCGACTGAATCAAGCAGGCGAGCTTTCTCTGCTGCAGA 100
 |||||
 321 pAspArgLeuAsnArgMetAsnGlnAlaSerArgLeuPheLeuLeuGlnT 338
 101 CAGTTGAGCTGGGCTCTGGGATGGTGATGACATTGACAGTGGCTGCC 150
 |||||
 338 hrValGluGlnGlySerTrpValGlyMetValThrPheAspSerAlaAla 354
 151 CATGTACAAAGTGAATCATACAGATAAAGAGTGGCAGTGCAGGGACAC 200
 |||||
 355 TyrValGlnSerGluLeuLysGlnLeuAsnSerGlyAlaAspArgSpLe 371
 201 ACTGCCAAAGATTACTCGCAGCAGCTTCAGGAGGAGGACGTCATCTGCA 250
 |||||
 371 uLeuIleLysHisLeuProThrValSerAlaGlyGlyThrSerIleCysS 388
 251 GC 252
 ||
 388 er 388

seq_name: sp_rodent:Q9D726

seq_documentation_block:
 ID Q9D726 PRELIMINARY; PRT; 913 AA.
 AC Q9D726;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CHANNEL CLCA3.
 GN CLCA3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=STOMACH;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nardone P., King B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK008659; BAB25815.1; -.
DR MGD; MGI:1346342; Clca3.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
SQ SEQUENCE 913 AA; 100111 MW; 4918584B06D9A89 CRC64;

alignment_scores:
Quality: 330.00 Length: 84
Ratio: 4.231 Gaps: 0
Percent Similarity: 92.857 Percent Identity: 77.381

alignment_block:

US-09-049-696-6 x Q9D726 ..

Align seg 1/1 to: Q9D726 from: 1 to: 913

1 CAAGAATTCTGTGTTAGTCCTTGACAAATCTGGAAGCATGCGACTGG 50
|||||
305 GlnArgIleValCysLeuValLeuAspLysSerGlySerMetLeuAsnAs 321
51 TAACCGCCTCAATCGACTGAATCAAGCAGCCAGCTTTCTCTCTGCAGA 100
|||||
321 pAspArgLeuAsnArgMetAsnGlnAlaSerArgLeuPheLeuGlnT 338
101 CAGTTGAGCTGGGCTGGTGGATGTGACATTGTGACATTTGCTGCTGCC 150
|||||
338 hrValGluGlnGlySerTrpValGlyMetValThrPheAspSerAlaAla 354
151 CATGTACAAAGTAGTACATACAGATAAAGAGTGGCAGTGCAGGACAC 200
|||||
355 TyrValGlnSerGluLeuLysGlnLeuAsnSerGlyAlaAspArgAspLe 371
201 ACTGCCAAAAGATTACTCGCAGCAGCTTCAGGAGGACGCTCCATCGCA 250
|||||
371 uLeuIleLysHisLeuProThrValSerAlaGlyGlyThrSerIleCys 388

251 GC 252

388 er 388

seq_name: sp_human:Q9UNF7

seq_documentation_block:

ID Q9UNF7 PRELIMINARY; PRT; 917 AA.

AC Q9UNF7
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 2.
GN CACCC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=COLON;
RX MEDLINE=99364503; PubMed=10437792;
RA Agnel M., Vermet T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
RT chloride channel (CaCC) family predominantly expressed in the
RT digestive tract and trachea."
RL FEBS Lett. 455:295-301(1999).

DR EMBL; AF127035; AAD48398.1; -.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00092; vwa; 1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 917 AA; 101153 MW; EA01030D781BEB95 CRC64;

alignment_scores:
Quality: 268.00 Length: 84
Ratio: 3.775 Gaps: 0
Percent Similarity: 84.524 Percent Identity: 66.667

alignment_block:

US-09-049-696-6 x Q9UNF7 ..

Align seg 1/1 to: Q9UNF7 from: 1 to: 917

1 CAAGAATTCTGTGTTAGTCCTTGACAAATCTGGAAGCATGCGACTGG 50
|||||
304 GlnArgIleValCysLeuValLeuAspLysSerGlySerMetGlyGlyLy 320
51 TAACCGCCTCAATCGACTGAATCAAGCAGCCAGCTTTCTCTCTGCAGA 100
|||||
320 sAspArgLeuAsnArgMetAsnGlnAlaLysHisPheLeuGlnT 337
101 CAGTTGAGCTGGGCTGGTGGATGTGACATTGTGACATTTGACAGTGCCTGCC 150
|||||
337 hrValGluGlnGlySerTrpValGlyMetValHisPheAspSerThrAla 353
151 CATGTACAAAGTAGTACATACAGATAAAGAGTGGCAGTGCAGGACAC 200
|||||
354 ThrIleValAsnLysLeuIleGlnIleLysSerSerAspGluArgAsnTh 370
201 ACTGCCAAAAGATTACTCGCAGCAGCTTCAGGAGGACGCTCCATCGCA 250
|||||
370 rLeuMetAlaGlyLeuProThrTyrProLeuGlyGlyThrSerIleCys 387

251 GC 252

387 er 387

seq_name: sp_mammal:Q92822

seq_documentation_block:

ID Q92822 PRELIMINARY; PRT; 109 AA.

AC Q92822;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-1997 (TrEMBLrel. 03, Last annotation update)
DE UNKNOWN PROTEIN (FRAGMENT).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=96327607; PubMed=8672129;
RA Winteroe A.K., Fredholm M., Davies W.;
RT "Evaluation and characterization of a porcine small intestine cDNA
RT library: analysis of 839 clones."
RL Mamm. Genome 7:509-517(1996).
DR EMBL; F15082; CAA23338.1; -.
FT NON_TER 1
FT NON_TER 109
SQ SEQUENCE 109 AA; 12048 MW; 67FCD61718B151CD CRC64;

alignment_scores:

Quality: 258.50 Length: 74

Ratio: 3.858 Gaps: 1

Percent Similarity: 90.541 Percent Identity: 72.973

alignment_block:

US-09-049-696-6 x Q29282

Align seg 1/1 to: Q29282 from: 1 to: 109

4 AGAATTGTGTGTAGTCCCTGACAAATCTGGAAGCATGGCGACTGTGTA 53
|||||
1 ArgileValCysLeuValLeuAspLysSerGlySerMetThrValGlyG 17
|||||
54 CGCGCTCAATCGACTGAATCAAGCAGCCAGCTTTCTGCTGCGACAG 103
|||||
17 yArgLeuLysArgLeuAsnGlnAlaGlyLysLeuPheLeuGlnThrV 34
|||||
104 TTGAGCTGGGTCTCGTGGTGGATGCTGACATTTGACAGTGTGCCCAT 153
|||||
34 alGluGlnGlyAlaTrpValGlyMetValAlaPheAspSerAlaAla 50
|||||
154 GTACAAAGTGAATCATACAGATAAAGTGGCAGTGCAGGAGGACACT 203
|||||
51 ValysSerGluLeuValGlnIleAsnSerAlaAlaAsp.....** 64

204 CGCCAAAGATTACCTGCACCA 225

*ArgGlnCysLeuProAlaAla 71

seq_name: sp_rodent:O88860

seq_documentation_block:

ID O88860 PRELIMINARY; PRT; 901 AA.
AC O88860;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CHLORIDE CHANNEL CACC.
GN CLCAL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RA ROMLO L., Musante L., Cinti R., Moran O., Seri M., Galletta L.J.V.;
RT "Characterization of a murine gene homologous to the bovine CaCC
RT Chloride channel."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF052746; AAC35003.1;
DR MGD; MGI:1316732; Clcal.
DR InterPro: IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 901 AA; 100038 MW; 13DF068BF3337AAB CRC64;

alignment_scores:

Quality: 248.00 Length: 83
Ratio: 3.647 Gaps: 0
Percent Similarity: 81.928 Percent Identity: 61.446

alignment_block:

US-09-049-696-6 x O88860

Align seg 1/1 to: O88860 from: 1 to: 901

1 CAAAGAATTCGTGTAGTCTTACAAATCTGGAAGCATGGCGACTGG 50
|||||
306 ArgArgValValCysLeuValLeuAspLysSerGlySerMetAspLysG 322
|||||
51 TAACCGCTCAATCGACTGAATCAAGCAGCCAGCTTTCTGCTGCGAGA 100
|||||
322 uAspArgLeuIleArgMetAsnGlnAlaGluLeuTyrlLeuThrGlnI 339
|||||
101 CAGTTGAGCTGGGTCTCGTGGTGGATTTGACATTTGACAGTGTGCC 150

|||||
339 leValGluLysGluSerMetValGlyLeuValThrPheAspSerAlaAla 355
|||||
151 CATGTACAAAGTGAATCATACAGATAAAGTGGCAGTGCAGGAGCAC 200
|||||
356 HisIleGlnAsnTyrLeuIleLysIleThrSerSerAspTyrGlnLy 372
|||||
201 ACTCGCCAAAGATTACCTGCAGCAGCTTCAGGAGGAGGACGTCATCTGC 249
|||||
372 sileThrAlaAsnLeuProGlnGlnAlaSerGlyGlyThrSerIleCys 388

seq_name: sp_rodent:Q9QX15

seq_documentation_block:

ID Q9QX15 PRELIMINARY; PRT; 902 AA.
AC Q9QX15;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CALCIUM-SENSITIVE CHLORIDE CONDUCTANCE PROTEIN-1.
GN CLCAL OR MCLCAL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RA Gaudhi R., Elble R.C., Gruber A.D., Schreier K.D., Ji H.-L.,
RA Fueller C.M., Pauli B.U.;
RT "Molecular and functional characterization of a calcium-sensitive
RT Chloride channel from mouse lung."
RL J. Biol. Chem. 273:32096-32101(1998).
DR EMBL; AF047838; AAC79982.1;
DR MGD; MGI:1316732; Clcal.
DR InterPro: IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 902 AA; 100107 MW; D9A915E698EA2D85 CRC64;

alignment_scores:

Quality: 248.00 Length: 83
Ratio: 3.647 Gaps: 0
Percent Similarity: 81.928 Percent Identity: 61.446

alignment_block:

US-09-049-696-6 x Q9QX15

Align seg 1/1 to: Q9QX15 from: 1 to: 902

1 CAAAGAATTCGTGTAGTCTTACAAATCTGGAAGCATGGCGACTGG 50
|||||
306 ArgArgValValCysLeuValLeuAspLysSerGlySerMetAspLysG 322
|||||
51 TAACCGCTCAATCGACTGAATCAAGCAGCCAGCTTTCTGCTGCGAGA 100
|||||
322 uAspArgLeuIleArgMetAsnGlnAlaGluLeuTyrlLeuThrGlnI 339
|||||
101 CAGTTGAGCTGGGTCTCGTGGTGGATTTGACATTTGACAGTGTGCC 150
|||||
339 leValGluLysGluSerMetValGlyLeuValThrPheAspSerAlaAla 355
|||||
151 CATGTACAAAGTGAATCATACAGATAAAGTGGCAGTGCAGGAGCAC 200
|||||
356 HisIleGlnAsnTyrLeuIleLysIleThrSerSerAspTyrGlnLy 372
|||||
201 ACTCGCCAAAGATTACCTGCAGCAGCTTCAGGAGGAGGACGTCATCTGC 249
|||||
372 sileThrAlaAsnLeuProGlnGlnAlaSerGlyGlyThrSerIleCys 388

seq_name: sp_rodent:Q9R070

```
seq_documentation_block:
ID Q9R070 PRELIMINARY; PRT; 902 AA.
AC Q9R070;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CA(2+)-SENSITIVE CHLORIDE CHANNEL 2.
GN CLCA2 OR CACC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20012773; PubMed=10544033;
TX Lee D., Ha S., Kho Y., Kim J., Cho K., Baik M., Choi Y.;
RT "Induction of mouse Ca(2+)-sensitive chloride channel 2 gene during
involvement of mammary gland."
RL Biochem. Biophys. Res. Commun. 264:933-937(1999).
DR EMBL: AF108501; AAF12731.1; --
DR MGI:1931471; Clca2.
DR InterPro; IPR002035; VFWA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VFWA; 1.
SQ SEQUENCE 902 AA; 99866 MW; 0FB746CF5FB3D07F CRC64;

alignment_scores:
Quality: 245.00 Length: 83
Ratio: 3.603 Gaps: 0
Percent Similarity: 81.928 Percent Identity: 60.241

alignment_block:
US-09-049-696-6 x Q9R070
Align seg 1/1 to: Q9R070 from: 1 to: 902
1 CAAGAAATGTGTGTTAGTCCTTGACAAATCTGGAAGCATGCGACTGG 50
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
306 ArgArgValValCysLeuValLeuAspLysSerGlySerMetAspLysG1 322
51 TAACCGCCTCAATCGACTGAATCAAGCAGCGCTTCTGCTGCGACA 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 uAspArgLeuLeuArgMetAsnGlnAlaGluLeuThrLeuThrGlnI 339
101 CAGTTGAGCTGGGCTGCTGCTGGATGGTGCATTTGACAGTGTGCC 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 leValGluLysGluSerMetValGlyLeuValThrPheAspSerAlaAla 355
151 CATGTACAAAGTGAATCATACAGATAAAGTGGCAGTGCAGGAGCAC 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356 HisIleGlnAsnTyrLeuIleLysIleThrSerSerAspTyrGlnLy 372
201 ACTGCCAAAGATTACCTGCAGCAGCTTCAGGAGGAGCTCCATGCG 249
:::|||||:|||||:|||||:|||||:|||||:|||||:
372 sileThrAlaAsnLeuProGlnGlnAlaThrGlyGlyThrSerIleCys 388
seq_name: sp_mammal:018742

seq_documentation_block:
ID Q13742 PRELIMINARY; PRT; 794 AA.
AC Q13742;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Ruminantia; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
TX Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001262; AAB86530.1; --
DR InterPro; IPR002035; VFWA.
DR PROSITE; PS50234; VFWA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 794 AA; 88509 MW; B695E7256FC2C632 CRC64;

seq_documentation_block:
ID Q9EQR4 PRELIMINARY; PRT; 902 AA.
AC Q9EQR4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ENDOTHELIAL CHLORIDE CHANNEL.
GN CLCA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```
alignment_scores:
  Quality: 233.00      Length: 83
  Ratio: 3.236         Gaps: 0
  Percent Similarity: 86.747  Percent Identity: 54.217

alignment_block:
US-09-049-696-6 x 018742
..
Align seg 1/1 to: 018742 from: 1 to: 794

1 CAAGAATTGTGTTAGTCTTGTGACAAATCTGGAAGCATGGGACTGG 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
307 GlnArgValValCysLeuValLeuAspLysSerGlySerMetSerAlaG1 323

51 TAACCGCTCAATCGACTGAATCAAGCAGCGCCAGCTTTCTGCTGCAGA 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
323 uAspArgLeuPheGlnMetAsnGlnAlaGluLeuTyrLeuIleGlnV 340

101 CAGTTGAGCTGGGTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
340 aliLeGluLysGlySerLeuValGlyMetValThrPheAspSerValAla 356

151 CATGTACAAAGTGAACATACATACATAAAGTGGCAGTGACAGGGAC 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
357 GluIleGlnAsnHisLeuThrArgIleThrAspAsnValTyrGlnLy 373

201 ACTGCCCAAAAGATTACTCGCAGCAGCTTCAGGAGGAGGAGTCCAT 249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
373 sileThrAlaLysLeuProGlnValAlaAsnGlyGlyThrSerIleCys 389

seq_name: sp_mammal:018741
seq_documentation_block:
ID 018741 PRELIMINARY; PRT; 905 AA.
AC 018741;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001261; AAB86529.1;
DR InterPro; IPR002035; VWFA.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00041; fn3; 1.
DR PROSITE; PS00234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR SEQUENCE 905 AA; 101005 MW; D4CA6E7919115E88 CRC64;

alignment_scores:
  Quality: 233.00      Length: 83
  Ratio: 3.236         Gaps: 0
  Percent Similarity: 86.747  Percent Identity: 54.217

alignment_block:
US-09-049-696-6 x 018741
..
Align seg 1/1 to: 018741 from: 1 to: 905

1 CAAGAATTGTGTTAGTCTTGTGACAAATCTGGAAGCATGGGACTGG 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
307 GlnArgValValCysLeuValLeuAspLysSerGlySerMetSerAlaG1 323

51 TAACCGCTCAATCGACTGAATCAAGCAGCGCCAGCTTTCTGCTGCAGA 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
323 uAspArgLeuPheGlnMetAsnGlnAlaGluLeuTyrLeuIleGlnV 340

101 CAGTTGAGCTGGGTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
340 aliLeGluLysGlySerLeuValGlyMetValThrPheAspSerValAla 356

151 CATGTACAAAGTGAACATACATACATAAAGTGGCAGTGACAGGGAC 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
357 GluIleGlnAsnHisLeuThrArgIleThrAspAsnValTyrGlnLy 373

201 ACTGCCCAAAAGATTACTCGCAGCAGCTTCAGGAGGAGGAGTCCAT 249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
373 sileThrAlaLysLeuProGlnValAlaAsnGlyGlyThrSerIleCys 389

seq_name: sp_mammal:018743
seq_documentation_block:
ID 018743 PRELIMINARY; PRT; 820 AA.
AC 018743;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001263; AAB86531.1;
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR SEQUENCE 820 AA; 91464 MW; 4E01330AEF5A1494 CRC64;

alignment_scores:
  Quality: 233.00      Length: 83
  Ratio: 3.236         Gaps: 0
  Percent Similarity: 86.747  Percent Identity: 54.217

alignment_block:
US-09-049-696-6 x 018743
..
Align seg 1/1 to: 018743 from: 1 to: 820

1 CAAGAATTGTGTTAGTCTTGTGACAAATCTGGAAGCATGGGACTGG 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
307 GlnArgValValCysLeuValLeuAspLysSerGlySerMetSerAlaG1 323

51 TAACCGCTCAATCGACTGAATCAAGCAGCGCCAGCTTTCTGCTGCAGA 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

This Page Blank (uspto)

OM of: US-09-049-696-6 to: SwissProt_39:* out_format : pfs

Date: Mar 30, 2002 2:51 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+g2p.model -DEV=xlp
-o=/cgn2_1/uspt0_spo01/US09049696/runat_28032002_145238_2085/app_query.fasta_1.12579
-DB=SwissProt_39 -QFMT=fastan -SUFFIX=isp -GAPOP=12.000
-CAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1_165 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-049-696-6
Query length: 252
Database: SwissProt_39:*
Database sequences: 100059
Database length: 36564827
Search time (sec): 306.030000
```

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
SwissProt_39:ECIC_BOVIN	+	236.00	481.29	4.3e-20	903	! P54281 bos taurus (bovine).
SwissProt_39:Y103_HUMAN	+	87.00	168.66	0.0241	420	! Q55874 synecocystis sp. (st
SwissProt_39:FATH_HUMAN	+	71.50	115.10	2.12	4590	! Q14517 homo sapiens (human).
SwissProt_39:Y550_MYCTU	+	66.50	121.31	6.98	629	! O05809 mycobacterium tubercu
SwissProt_39:YQ12_CAHEL	+	65.50	111.48	9.98	1551	! Q09449 caenorhabditis elega
SwissProt_39:YGR1_YEAST	+	65.00	117.95	10.56	640	! P07911 homo sapiens (human).
SwissProt_39:ATX9_YEAST	+	64.50	111.88	12.79	1151	! P24814 saccharomyces cervisi
SwissProt_39:ATX9_YEAST	+	64.00	110.94	14.65	1133	! Q95050 tetrahymena thermoph
SwissProt_39:FLAG_BACSU	+	63.50	122.27	14.70	264	! P23446 bacillus subtilis. fl
SwissProt_39:CRAR_HUMAN	+	63.00	112.91	18.44	699	! Q48740 h complement-activati
SwissProt_39:YC81_CAHEL	+	63.00	107.69	19.51	1291	! Q19981 caenorhabditis elega
SwissProt_39:APXL_HUMAN	+	63.00	105.78	19.92	1616	! Q13796 homo sapiens (human).
SwissProt_39:CH60_COWRU	+	62.50	113.87	20.70	551	! P48213 cowdria ruminantium.
SwissProt_39:LIP_RHIO	+	62.00	115.69	23.02	392	! P21811 rhizopus oryzae (rhiz
SwissProt_39:YC03_KLEPN	+	62.00	113.55	23.56	504	! Q48449 klebsiella pneumoniae
SwissProt_39:THRB_MOUSE	+	62.00	111.82	24.00	618	! P19221 mus musculus (mouse).
SwissProt_39:TSPL_HUMAN	+	62.00	106.39	25.45	1170	! P07996 homo sapiens (human).
SwissProt_39:TSPL_MOUSE	+	62.00	106.39	25.45	1170	! P35441 mus musculus (mouse).
SwissProt_39:VMT2_HUMAN	+	61.50	112.32	27.07	514	! Q01827 rattus norvegicus (rat).
SwissProt_39:VMT2_RAT	+	61.50	112.30	27.08	515	! Q01827 rattus norvegicus (rat).
SwissProt_39:THYG_HUMAN	+	61.50	97.98	31.60	2768	! P01266 homo sapiens (human).
SwissProt_39:ECL_ARATH	+	61.00	126.66	26.30	84	! P33746 arabidopsis thaliana (m
SwissProt_39:CO7_HUMAN	+	61.00	107.03	32.50	843	! P10643 homo sapiens (human).
SwissProt_39:FLIM_CAUCR	+	60.50	112.90	34.60	373	! P34009 caulobacter crescentu
SwissProt_39:YINC_PSEFL	+	60.50	109.28	35.98	571	! P23031 pseudomonas fluoresce
SwissProt_39:MYIF_MOUSE	+	60.50	103.71	38.21	1099	! P70248 mus musculus (mouse).
SwissProt_39:XDH_DROME	+	60.50	102.05	38.89	1335	! P10351 drosophila melanogast
SwissProt_39:XDH_DROSU	+	60.50	101.99	38.92	1344	! P19171 drosophila subobscura
SwissProt_39:NUAM_ANOQU	+	60.00	110.27	40.37	448	! P33511 anopheles quadrimacul
SwissProt_39:ATKA_ECOLI	+	60.00	108.42	41.18	557	! P33959 escherichia coli. pot
SwissProt_39:YPR2_MOUSE	+	60.00	89.20	50.66	5327	! Q9qxx0 mus musculus (mouse).
SwissProt_39:YPR2_ECOLI	+	59.50	119.67	41.37	131	! P10018 escherichia coli. hyp
SwissProt_39:ENV_GALV	+	59.50	105.81	48.03	667	! P21415 gibbon ape leukemia vir
SwissProt_39:HGL_HUMAN	+	59.50	105.27	48.32	711	! P26927 homo sapiens (human).
SwissProt_39:IREL_RABIT	+	59.50	103.37	49.32	889	! Q01059 oryctolagus cuniculus
SwissProt_39:Y1379_HUMAN	+	59.50	101.88	50.11	1059	! Q15084 homo sapiens (human).
SwissProt_39:NUAM_ANOAR	+	59.00	114.15	49.79	221	! P51898 anopheles arabiensis
SwissProt_39:UPPS_MYCTU	+	59.00	111.66	51.15	296	! O05837 mycobacterium tubercu
SwissProt_39:Y456_MYCPN	+	59.00	110.35	51.87	345	! P75121 mycoplasma pneumoniae
SwissProt_39:NUAM_ANOGA	+	59.00	108.15	53.12	447	! P34852 anopheles gambiae (ad

alignment_scores:

Quality: 236.00 Length: 83

```
SwissProt_39:PD25_MYCTU + 59.00 105.89 54.43 583 ! Q05086 mycobacterium tube
SwissProt_39:CTRI_HUMAN - 59.00 105.24 54.81 629 ! P30825 homo sapiens (huma
SwissProt_39:LMAL_MOUSE - 59.00 91.71 63.42 3084 ! P19137 mus musculus (mou
SwissProt_39:TCPL_VIBCH - 58.50 116.66 54.96 145 ! P29489 vibrio cholerae. t
SwissProt_39:KDUL_ERWCH - 58.50 111.12 58.34 278 ! Q05529 erwinia chrysanth
seq_name: SwissProt_39:ECIC_BOVIN
seq_documentation_block:
ID ECIC_BOVIN STANDARD; PRT; 903 AA.
AC P54281;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE EFITHIAL CHLORIDE CHANNEL PROTEIN (CALCIUM-ACTIVATED CHLORIDE
DE CHANNEL).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos..
OX NCBI_TaxID=9913;
RN [1];
RP SEQUENCE FROM N.A.
RC TISSUE=Trachea;
RX MEDLINE=96125078; PubMed=8537359;
RA Cunningham S.A., Awayda M.S., Buben J.K., Ismailov I.I.,
RA Arrate M.P., Berdiev B.K., Benos D.J., Fuller C.M.;
RT "Cloning of an epithelial chloride channel from bovine trachea.";
RL J. Biol. Chem. 270:31016-31026(1995).
CC !- FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE
CC SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME;
CC MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND
CC TRANSEPITHelial TRANSPORT.
CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC !- TISSUE SPECIFICITY: TRACHEA.
CC !- PTM: PHOSPHORYLATED BY CAM-KINASE II.
CC !- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U36445; AAC48511.1; -
CC InterPro: IPR002035; VWFA.
CC SMART: SM00327; VWFA; 1.
CC PROSITE: PS50234; VWFA; 1.
CC Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
CC Calcium channel; Phosphorylation; Glycoprotein.
CC TRANSMEM 7 27
CC TRANSMEM 331 351
CC TRANSMEM 617 637
CC TRANSMEM 883 903
CC DOMAIN 308 476
CC N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 75 75
CC CARBOHYD 278 278
CC CARBOHYD 360 360
CC CARBOHYD 372 372
CC CARBOHYD 504 504
CC CARBOHYD 515 515
CC CARBOHYD 688 688
CC CARBOHYD 811 811
CC CARBOHYD 816 816
CC CARBOHYD 842 842
CC CARBOHYD 857 857
CC SEQUENCE 903 AA; 100305 MW; 9742AB1590908AEC CRC64;
```


FT	CARBOHYD	1751	1751	N-LINKED	(GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	1867	1867	N-LINKED	(GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	1905	1905	N-LINKED	(GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	1943	1943	N-LINKED	(GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	1994	1994	N-LINKED	(GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	2328	2328	N-LINKED	(GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	2467	2467	N-LINKED	(GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	3326	3326	N-LINKED	(GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	3424	3424	N-LINKED	(GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	3446	3446	N-LINKED	(GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	3615	3615	N-LINKED	(GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	3642	3642	N-LINKED	(GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	3718	3718	N-LINKED	(GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	4154	4154	N-LINKED	(GLCNAC..)	(POTENTIAL)
SO	SEQUENCE	4590 AA;	506273 MW;	04483CCD000E60A7	CRC64;	

Alignment scores:		Quality:	71.50	Length:	96
Ratio:		1.625		Gaps:	4
Percent Similarity:		45.833		Percent Identity:	27.083

Alignment block:		US-09-049-696-6 x FATH_HUMAN	..
Align seg 1/1 to: FATH_HUMAN		from: 1	to: 4590
34	GGAAGCATGGCGACTGCTTAACCGCTCAATCGACTGAATCAACGAGGC...	81	
1620	GlySerIleLysThrAlaLysGluLeuAspArgSerAsnGlnAlaGluTy	1636	
81	81	
1636	rAspLeuMetValLysAlaThrAspLysGlySerProMetSerGluI	1653	
82CAGCTTTTC.....	90	
1653	leThrSerValArgIlePheValThrIleAlaAspAsnAlaSerProLys	1669	
91CTGCTGCACACAGCTTCAGCTGGG	113	
1670	PheThrSerLysGluTySerValGluLeuSerGluThrValSerIleG	1686	
114	GTCCTGGTGGATGTGACATTTGACAGTGTGCCCATGTACAAAGTG	163	
1686	ySerPheValGlyMetValThr.....AlaHisSerGlnSerS	1699	
164	AACTCATA...CAGATAACAGTGGCAGTCACAGGAGC	198	
1699	erValValTyGluIleLysAspGlyAsnThrGlyAsp	1711	

seq_name: SwissProt_39:YS50_MYCTU	
seq_documentation_block:	
ID	YS50_MYCTU STANDARD; PRT; 629 AA.
AC	C05809;
DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DE	20-AUG-2001 (Rel. 40, Last annotation update)
DE	HYPOTHEICAL 67.0 KDA PROTEIN RV2850C.
GN	RV2850C OR MT2916 OR MTCY24AI.07.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacter
OX	NCBI_TaxID=1773;
RN	SEQUENCE FROM N.A.
RP	STRAIN=H37RV;
RC	MDLLINE=98295987; PubMed=9634230;
EX	Co6e S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harr
RA	Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroy

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87177970; PubMed=34531112;
 RA Pennica D., Kohr W.J., Kuang W.-J., Glaister D., Aggarwal B.B.,
 RA Chen E.Y., Goeddel D.V.;
 RT "Identification of human uromodulin as the Tamm-Horsfall urinary
 RT glycoprotein.";
 RL Science 236:83-88(1987).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=87319675; PubMed=3498215;
 RA Hession C., Decker J.M., Sherblom A.P., Kumar S., Yue C.C.,
 RA Mattaliano R.J., Tizard R.J., Kawashima E., Schweissner U.,
 RA Heletsky S., Chow E.P., Burne C.A., Shaw A., Muchmore A.V.;
 RT "Uromodulin (Tamm-Horsfall glycoprotein): a renal ligand for
 RT lymphokines.";
 RL Science 237:1479-1484(1987).
 RN [3]
 RP GPI-ANCHOR.
 RX MEDLINE=91065873; PubMed=2249987;
 RA Rindler M.J., Naik S.S., Li N., Hoops T.C., Peraldi M.-N.;
 RT "Uromodulin (Tamm-Horsfall glycoprotein/uromucoid) is a
 RT phosphatidylinositol-linked membrane protein.";
 RL J. Biol. Chem. 265:20784-20789(1990).
 CC -!- FUNCTION: NOT KNOWN. MAY PLAY A ROLE IN REGULATING THE CIRCULATING
 CC ACTIVITY OF CYTOKINES AS IT BINDS TO IL-1, IL-2 AND TNF WITH HIGH
 CC AFFINITY.
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR,
 CC THEN CLEAVED TO PRODUCE A SOLUBLE FORM WHICH IS SECRETED IN
 CC URINE.
 CC -!- TISSUE SPECIFICITY: SYNTHESIZED BY THE KIDNEYS AND IS THE MOST
 CC ABUNDANT PROTEIN IN NORMAL HUMAN URINE.
 CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 ZP DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M15881; AAA36798.1; -;
 DR EMBL; M17778; AAA36799.1; -;
 DR PIR; A30452; A30452.
 DR HSP; P07204; IEGR.
 DR GlycoSuiteDB; P07911; -;
 DR MIM; 191845; -;
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001507; zona_pellucida.
 DR Pfam; PF00008; EGF; 3.
 DR Pfam; PF00100; zona_pellucida; 1.
 DR PRINTS; PR00023; ZPELLUCIDA.
 DR SMART; SM00179; EGF_CA; 2.
 DR SMART; SM00001; EGF_like; 1.
 DR SMART; SM00241; ZP; 1.
 DR PROSITE; PS00682; ZP_DOMAIN; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 2.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 3.
 DR PROSITE; PS01187; EGF_CA; 2.
 DR GlycoProtein; Signal; Membrane; GPI-anchor; EGF-like domain.
 KW SIGNAL 1 24
 FT CHAIN 25 640 UROMODULIN.
 FT DOMAIN 28 64 EGF-LIKE 1.
 FT DOMAIN 65 107 EGF-LIKE 2.
 FT DOMAIN 108 149 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
 FT

FT DOMAIN 334 585 ZP.
 FT DISULFID 32 41 BY SIMILARITY.
 FT DISULFID 35 50 BY SIMILARITY.
 FT DISULFID 52 63 BY SIMILARITY.
 FT DISULFID 69 83 BY SIMILARITY.
 FT DISULFID 77 92 BY SIMILARITY.
 FT DISULFID 94 106 BY SIMILARITY.
 FT DISULFID 112 126 BY SIMILARITY.
 FT DISULFID 120 135 BY SIMILARITY.
 FT DISULFID 137 148 BY SIMILARITY.
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 76 76 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 80 80 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 232 232 N-LINKED (GLCNAC. .).
 FT CARBOHYD 275 275 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 396 396 N-LINKED (GLCNAC. .).
 FT CONFLICT 565 565 H -> D (IN REF. 2).
 SQ SEQUENCE 640 AA; 69760 MW; D26A07A76353AE48 CRC64;
 alignment_scores:
 Quality: 65.00 Length: 53
 Ratio: 2.241 Gaps: 3
 Percent Similarity: 54.717 Percent Identity: 32.075
 alignment_block:
 US-09-049-696-6/rev x UROM_HUMAN ..
 Align seg 1/1 to: UROM_HUMAN from: 1 to: 640
 190 CACTGCCACGTGTTATCTGTATGATGTCACTTGTA..... 155
 |||||..... ||||| |||||
 115 HiscysHisAlaLeuAlaThrCysValAsnValValGlySerTyrLeuCy 135
 154CATGGCAGCAGCTGTCAATGTCAC 130
 135 svalcysProAlaGlyTyrArgGlyAspGlyTyrHisCysGluCysSerP 152
 |||||..... |||||
 129 CATCCCAACAGCAGCCAGCTCAACTGTCTGCAGCAG...GAAAGCT 83
 |||||..... ||||| |||||
 152 roGly.SerCysGlyProGlyLeuAspCysValProGlyGlyAspAlaLe 168
 82 GGCTGTC 76
 |||||
 168 uValCys 170
 seq_name: SwissProt_39:GRRL_YEAST
 seq_documentation_block:
 ID GRRL_YEAST STANDARD; PRT; 1151 AA.
 AC P24814;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE GRRL PROTEIN.
 GN GRRL OR COT2 OR CAT80 OR YJR090C OR J1885.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92017785; PubMed=1922034;
 RA Flick J.S., Johnston M.;
 RT "GRRL of Saccharomyces cerevisiae is required for glucose repression
 RT and encodes a protein with leucine-rich repeats.";
 RL Mol. Cell. Biol. 11:5101-5112(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=96437976; PubMed=8840504;
 RA Huang M.-E., Manus V., Chuat J.-C., Gallibert F.;

01-FEB-1996 (Rel. 33, Last sequence update)
20-AUG-2001 (Rel. 40, Last annotation update)
DE COMPLEMENT-ACTIVATING COMPONENT OF RA-REACTIVE FACTOR PRECURSOR
DE (EC 3.4.21.-) (RA-REACTIVE FACTOR SERINE PROTEASE P100) (RARF)
DE (MANNA-BINDING LECTIN SERINE PROTEASE 1) (MANNOSE-BINDING PROTEIN
DE ASSOCIATED SERINE PROTEASE) (MASP-1).
GN MASP1 OR CRARF OR CRARF1 OR PRSS5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=9405062; PubMed=8240317;
RA Takada F., Takayama Y., Hatusue H., Kawakami M.;
RT "A new member of the C1s family of complement proteins found in a
RL bactericidal factor, Ra-reactive factor, in human serum.";
RL Biochem. Biophys. Res. Commun. 196:1003-1009(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal liver;
RX MEDLINE=94289349; PubMed=8018603;
RA Sato T., Endo Y., Matsushita M., Fujita T.;
RT "Molecular characterization of a novel serine protease involved in
RL activation of the complement system by mannose-binding protein.";
RL Int. Immunol. 6:665-669(1994).
CC -1- FUNCTION: COMPONENT OF THE BACTERICIDAL RA-REACTIVE FACTOR RARE
CC WHICH SPECIFICALLY BINDS TO RA AND R2 POLYSACCHARIDES EXPRESSED BY
CC CERTAIN ENTEROBACTERIA. IT TRIGGERS THE ACTIVATION OF COMPLEMENT
CC CASCADE BY ACTIVATING THE C4 AND C2 COMPONENTS. IT ACTIVATES THE
CC C4 COMPONENT BY CLEAVING THE ALPHA-CHAIN OF C4.
CC -1- SUBUNIT: RARF CONSISTS OF A COMPLEMENT-ACTIVATING COMPONENT
CC (CRARF) AND A POLYSACCHARIDE-BINDING (MANNOSE-BINDING) COMPONENT.
CC CRARF IS A HETERODIMER OF A HEAVY (P70) AND A LIGHT CHAIN (P29)
CC LINKED BY A DISULFIDE BOND.
CC -1- DOMAIN: CRARF HAS A MODULE ORGANIZATION SIMILAR TO CIR AND C1S.
CC -1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D17525; BAA04477.1;
CC EMBL; D28593; BAA05928.1;
CC HSSP; P00736; LAPO.
CC MEMOPS; S01198;
CC MIM; 600521;
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR000859; CUB.
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR001881; EGF_Ca.
CC InterPro: IPR000436; Sush1_SCR_CCP.
CC InterPro: IPR001254; Trypsin.
CC Pfam; PF00431; CUB; 2.
CC Pfam; PF00084; sush1; 2.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00032; CCP; 2.
CC SMART; SM00042; CUB; 2.
CC SMART; SM00179; EGF_CA; 1.
CC SMART; SM00020; Tryp_SPC; 1.
CC PROSITE; PS00010; ASX_HYDROXYL; 1.
CC PROSITE; PS01180; CUB; 2.

DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Complement pathway; Serine protease; Protease;
KW Glycoprotein; Sush1; Repeat; Signal; EGF-like domain; Hydroxylation.
FT SIGNAL 1 19
FT CHAIN 20 699
FT FT
FT FT
FT CHAIN 20 448
FT CHAIN 449 699
FT CHAIN 20 138
FT DOMAIN 139 182
FT DOMAIN 185 297
FT DOMAIN 300 363
FT DOMAIN 366 433
FT DOMAIN 449 699
FT ACT_SITE 490 490
FT ACT_SITE 552 552
FT ACT_SITE 646 646
FT MOD_RES 159 159
FT DISULFID 73 91
FT DISULFID 143 157
FT DISULFID 153 166
FT DISULFID 168 181
FT DISULFID 185 212
FT DISULFID 242 260
FT DISULFID 301 349
FT DISULFID 329 362
FT DISULFID 367 414
FT DISULFID 397 432
FT DISULFID 436 572
FT DISULFID 614 631
FT DISULFID 642 672
FT CARBOHYD 49 49
FT CARBOHYD 178 178
FT CARBOHYD 385 385
FT CARBOHYD 407 407
FT CONFLICT 235 235
FT CONFLICT 285 285
FT CONFLICT 499 499
FT CONFLICT 543 543
FT CONFLICT 643 643
SQ SEQUENCE 699 AA; 79261 MW; 45DEC2EA6B40151 CRC64;
alignment_scores: Quality: 63.00 Length: 69
Ratio: 1.909 Gaps: 4
Percent Similarity: 47.826 Percent Identity: 33.333
alignment_block:
US-09-049-696-6 x CRAR_HUMAN ..
Align seg 1/1 to: CRAR_HUMAN from: 1 to: 699
25 GACAAATCTGGAGCATGGCGACTGTACCGCCTCAATCGACTGAATCA 74
645 AspSerGlyGlyProMetValThrLeuAsnArg.....GI 656
75 AGCAGGCCAGCTTTCTCTGCTGCAGACACTTGAGCTGGGGTCTCTGGGTG 124
656 uArgGlyGlnTrpTyrLeuValGlyThrVal.SerTrpGlyAspAspCys 672
125 GGATGTGACATTTTGACAGCTGTGCCCACTGACAAAGTCAACTACATACAG 174
673 Gly.....LysLysAspArgTyrGlyValTyrSerTyr... 683
175 ATAAACAGTGGCAGTGACAGGGACACACTCGCCAAAGATTACCTGCAGC 224
684IleHisAsnLysAspTrrPileGlnA 693

APXL_HUMAN STANDARD; PRT: 1616 AA.
 AC Q13796;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE APICAL-LIKE PROTEIN (APXL PROTEIN).
 GN APXL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retina;
 RX MEDLINE=95315933; PubMed=7795590;
 RA Schiaffino V.M., Bassi M.T., Rugarli E.I., Renieri A., Galli L.,
 RA Ballabio A.;
 RT "Cloning of a human homologue of the Xenopus laevis APX gene from the
 RT ocular albinism type 1 critical region.";
 RL Hum. Mol. Genet. 4:373-382(1995).
 RN [2]
 RP SEQUENCE OF 56-1616 FROM N.A.
 RA Shen Y., Gibbs R.A.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RC -!- TISSUE SPECIFICITY: ABUNDANT IN RETINA AND MELANOMA; ALSO IN
 CC BRAIN, PLACENTA, LUNG, KIDNEY, PANCREAS.
 CC -!- SIMILARITY: SOME, TO XENOPUS LAEVIS APICAL PROTEIN (APX).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X83543; CAA58534.1; -.
 DR EMBL; AC002365; AAC32592.1; -.
 DR HSSP; Q12959; 1PDR.
 DR MIM; 300103; -.
 DR InterPro: IPR000867; IGFBP.
 DR InterPro: IPR001478; PDZ.
 DR Pfam; PF00219; IGFBP; 1.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 FT SIMILAR 54 112
 FT TO MOUSE SYNTROPHIN-1 AND RABBIT
 FT DYSTROPHIN-ASSOCIATED PROTEIN.
 FT DOMAIN 150 153
 FT POLY-SER.
 FT DOMAIN 314 320
 FT POLY-PRO.
 FT DOMAIN 343 346
 FT POLY-ALA.
 FT DOMAIN 1065 1068
 FT POLY-PRO.
 SQ SEQUENCE 1616 AA; 176409 MW; 75240685BC0B60A2 CRC64;

alignment_scores:
 Quality: 63.00 Length: 78
 Ratio: 1.432 Gaps: 2
 Percent Similarity: 56.410 Percent Identity: 30.769

alignment_block:

US-09-049-696-6/rev x APXL_HUMAN

Align seg 1/1 to: APXL_HUMAN from: 1 to: 1616

252 GCTGACATGGACGCTCCCTGCTGAAGTGTGTCAGGTAATCTT..... 210

1077 AlathraspGlyAlaProAla.AspAlaProValGlyValLeuGlyArgp 1093

209TTGGCGAGTGTGCTCCCTGTCTACGTCGCCATGTTATCTGTATGAGT 165

1093 roPheProThrProSerProAlaSerLeuAspValTyrValAlaArgLeu 1109

121 GTTGGGATGTGACATTTGACAGTGTGTCGCCATGTACAAAGTAACTCAT 170

334 IleGlySerValasp...AsnSerCysAlaHisValGlnSerArgIleCy 349

1164 TCACCTTTGTACATGGGAGCAGCTGTCAAAATGTCAACATCCCAACCCAGGA 1115

1110 SerLeuHisSerProSerValPheSerSerAlaGlnProGlnAspTh 1126

114 CCCAGCTCAACTGTCTGTGACAGAGAAAGCTGGCTGTGTTGATTGAGTC 65

1126 rPolysAlaThrValCysGluArgGlySerGlnHisValSerGlyAspa 1143

64 GATTGAGCGGTTACACGATGCCATGCTTCCA 33
 1143 laSerArgProLeuProGluAlaLeuLeuPro 1153
 seq_name: SwissProt_39:CH60_COWRU
 seq_documentation_block:
 ID CH60_COWRU STANDARD; PRT: 551 AA.
 AC P48213;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE 60 KDA CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN).
 GN GROEL OR MOXA OR GROEL.
 OS Cowdria ruminantium.
 CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Ehrlichiae; Cowdria.
 CC RCHL_TaxID=779;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WELGEVONDEN;
 RX MEDLINE=96118689; PubMed=7496520;
 RA Lally N.C., Nicoll S., Paxton E.A., Cary C.M., Sumption K.J.;
 RA "The Cowdria ruminantium groE operon";
 RL Microbiology 141:2091-2100(1995).
 CC -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
 CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
 CC CONDITIONS (BY SIMILARITY).
 CC -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
 CC 7 SUBUNITS (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U13638; AAA93153.1; -.
 DR HSP; P06139; IGRLL.
 DR InterPro: IPR001844; Chaperonins_cpn60.
 DR InterPro: IPR002423; TCPL_cpn60.
 DR Pfam; PF00118; cpn60_TCPL; 1.
 DR PRINTS; PR00298; CHAPERONIN60.
 DR PRINTS; PR00304; TCOMPLEXTCP1.
 DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
 DR Chaperone; ATP-binding.
 SQ SEQUENCE 551 AA; 58829 MW; 936118790FD4E277 CRC64;

alignment_scores:

Quality: 62.50 Length: 49

Ratio: 1.953 Gaps: 4

Percent Similarity: 65.306 Percent Identity: 42.857

alignment_block:

US-09-049-696-6 x CH60_COWRU

Align seg 1/1 to: CH60_COWRU from: 1 to: 551

121 GTTGGGATGTGACATTTGACAGTGTGTCGCCATGTACAAAGTAACTCAT 170

334 IleGlySerValasp...AsnSerCysAlaHisValGlnSerArgIleCy 349

Tue Apr 2 09:40:11 2002

RT virulent strain Chetid. ",
 RL J. Bacteriol. 177:1788-1796(1995).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D21242; BAA04774.1; -
 CC Hypothetical protein.
 KW SEQUENCE 504 AA; 55782 MW; AD887595CFDFDFA8 CRC64;
 SQ

```

alignment_scores:      62.00      Length:      62
                       Ratio:    1.512      Gaps:      4
                       Percent Similarity: 35.484

alignment_block:
US-09-049-696-6 x YC03_KLEPN ..
..
Align seg 1/1 to: YC03_KLEPN from: 1 to: 504

69  GANTCAAGCAGGCCAGCTTTTCCTGCTGCAGACAGTTGAGCTGGGGTC... 116
   ::::  :::  |||||||  :::::  ||:::  |||||
446  GlnSerIleasnLysAlaPheProHisSerAspThrLeuLysGlyValcI 462

117  .CTGGGTGGGATGTGCACATTTTCAGATGCTGCCCATGTCAAAGTGAA 165
   ::::  :::::  |||  :::::  :::::  :::::  :::::  :::::
462  nLeuGlyTripsSerGlyaspValTyrGln.SerValArgLeuAsnThrSer 478

166  CTCATACAGATAAC... AGTGGCAGTGCAGGAGCACACACTCGCCAAAG 212
   |||  |||  |||  |||  |||  |||  |||  |||
479  LeuTrpTyrThrAsnAlaAsnAsnSerAspSeraspVal..... 492

213  ATTACCTGCAGCAGCTTTCAGGAGGAGCTGCATC 246
   ::|||:::  ::|||:::  ::|||
493  .....GlyAlaSerAlaGlyIleLeuIle 500

```

OM of: US-09-049-696-6 to: PIR_68:* out_format : pfs

Date: Mar 30, 2002 2:26 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+2np.model -DEV=xlp
-O=/cgn2_1/uspro_spool/us09049696/runat_28032002_145238_2031/app_query_fasta_1.12579
-DB=PIR_68 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -XGAPEXT=6.000
-GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THRF_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09049696 -CGN1_1_367 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-049-696-6

Query length: 252

Database: PIR_68*

Database sequences: 219241

Database length: 7617452

Search time (sec): 470.790000

score_list:

Sequence	Strtd Orig	ZScore	EScore	Len	Documentation
pir2:JG0168	+ 330.00	658.19	1.2e-29	913	! gob-5 protein - mouse
pir2:T02205	+ 233.00	458.71	1.6e-18	905	! Lu-ECAM-1 protein - bovine
pir2:T76691	+ 74.00	163.90	0.0922	420	! hypothetical protein - Synchoc
pir2:H81800	+ 77.00	138.34	2.88	356	! conserved hypothetical integral
pir2:D82493	+ 71.50	128.92	5.32	646	! conserved hypothetical protein
pir2:JC5953	+ 70.50	124.23	6.73	932	! inter-alpha-inhibitor H4p heavy
pir2:F84811	+ 70.00	125.37	7.87	689	! probable retroelement pol polyf
pir2:T40998	+ 69.00	125.43	10.48	513	! hypothetical protein SPC1450.1
pir2:S30400	+ 68.50	129.21	12.60	263	! hypothetical protein 1 - Strept
pir2:G85569	+ 68.00	122.78	13.56	557	! hypothetical protein kdpA [imp
pir2:A48569	+ 67.00	120.90	13.28	724	! antigen Em100 - Eimeria maxima
pir2:T29618	+ 67.00	120.23	17.94	452	! hypothetical protein D1014.1 -
pir2:E72453	+ 67.00	116.53	16.86	999	! probable cytochrome C-type blo
pir2:E70589	+ 66.50	118.82	19.95	629	! protoporphyrin IX magnesium che
pir2:G96651	+ 66.00	122.95	24.07	307	! protein T3P18.17 [imported] - A
pir2:T18941	+ 65.50	110.28	24.20	1551	! hypothetical protein C05C10.2
pir2:H81694	+ 65.00	127.47	33.65	123	! inclusion membrane protein B TC
pir1:A30452	+ 65.00	115.61	29.59	640	! uromodulin precursor - human
pir2:G83374	+ 64.50	118.47	35.22	373	! hypothetical protein PA0574 [im
pir2:A41529	+ 64.50	110.37	32.25	1151	! GRRI protein - yeast (Sacchar
pir2:T30302	+ 64.00	109.45	36.85	1133	! P-type ATPase - Tetrahymena th
pir2:G69622	+ 63.50	118.89	47.10	264	! flagellar hook protein flgE - E
pir2:G82599	+ 63.00	122.03	56.22	148	! hypothetical protein XF2118 [im
pir2:T51445	+ 63.00	118.83	54.30	231	! hypothetical protein F2G14.120
pir2:T29188	+ 63.00	114.34	51.73	431	! hypothetical protein C5C3.5 -
pir2:T47637	+ 63.00	111.11	49.94	676	! hypothetical protein T5N23.140
pir1:154763	+ 63.00	110.87	49.81	699	! Ra-reactive factor (EC 3.4.21.-
pir2:F83456	+ 63.00	109.91	49.29	799	! xanthine dehydrogenase PA1523
pir1:ORHUP	+ 63.00	108.79	48.70	933	! progesterone receptor form B -
pir2:T21694	+ 63.00	106.46	47.48	1291	! hypothetical protein F3C8.1 -
pir2:T37183	+ 63.00	104.84	46.66	1616	! gene APX1 protein - human
pir2:JC7500	+ 62.50	108.89	56.25	798	! qik protein - chicken
pir2:JT0604	+ 62.00	114.96	69.33	297	! triacylglycerol lipase (EC 3.1.1
pir2:JT01390	+ 62.00	112.97	67.84	392	! triacylglycerol lipase (EC 3.1.1
pir2:E82087	+ 62.00	111.59	66.83	475	! thymidine phosphorylase VC2349
pir1:A53827	+ 62.00	109.69	65.48	618	! thrombin (EC 3.4.21.5) B chain
pir1:TSHUP1	+ 62.00	105.11	62.30	1170	! thrombospondin 1 precursor - h
pir2:A40558	+ 62.00	105.11	62.30	1170	! thrombospondin 1 precursor - m
pir2:S62023	+ 61.50	119.24	83.79	142	! probable membrane protein YDR54
pir2:A49368	+ 61.50	109.99	75.79	514	! vesicle monoamine transporter 4

pir2:A45374 - 61.50 109.98 75.78 515 ! reserpine-sensitive vesicula
pir2:T07903 - 61.50 109.74 75.59 532 ! tubulin delta chain Uni3 - C
pir2:E69769 + 61.50 107.74 73.96 703 ! hypothetical protein ydan -
pir1:U1HU - 61.50 97.89 66.47 2767 ! thyroglobulin precursor, ma
seq_name: pir2:JG0168
seq_documentation_block:
gob-5 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
C:Accession: JG0168
R:Komiyama, T.; Tanigawa, Y.; Hirohashi, S.
Biochem. Biophys. Res. Commun. 255, 347-351, 1999
A:Title: Cloning and identification of the gene gob-5, which is expressed in intestine
A:Reference number: JG0168; MUID:99160866
A:Accession: JG0168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-913<KOM>
A:Cross-references: DDBJ:AB016592

alignment_scores:

Quality: 330.00 Length: 84

Ratio: 4.231 Gaps: 0

Percent Similarity: 92.857 Percent Identity: 77.381

alignment_block:

US-09-049-696-6 x JG0168 ..

Align seg 1/1 to: JG0168 from: 1 to: 913

1 CAAGAATTGCTTTAGTCTTGCACAAATCTGGAAGCATGGCAGTCGG 50
|||||
305 GlnArgileValCysLeuValLeuAspLysSerGlySerMetLeuAsn 321
51 TAACCGCTCAATCGACTGATCAAGCAGGCGCAGCTTTCTCTGTCGAGA 100
|||||
321 pAspArgLeuAsnArgMetAsnGlnAlaSerArgLeuPheLeuGlnT 338
101 CAGTTGAGCTGGGCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 150
|||||
338 hrValGluGlnGlySerTrpValGlyMetValThrPheAspSerAlaAla 354
151 CATGTACAAAGTGAACATCATACAGATAAACAGTGGCAGTGCAGGACAC 200
|||||
355 TyrValGlnSerGluLeuLysGlnLeuAsnSerGlyAlaAspArgAspLe 371
201 ACTGCCCAAGAGATTACTGCGCAGCAGCTTCAGGAGGACGTCCTCATCTGCA 250
|||||
371 uLeuileLysHisLeuProThrValSerAlaGlyThrSerileCysS 388
251 GC 252
388 er 388

seq_name: pir2:T02205

seq_documentation_block:

Lu-ECAM-1 protein - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 01-Dec-2000
C:Accession: T02205; T02152; T02171
R:Elble, R.C.; Widom, J.; Gruber, A.D.; Abdel-Ghany, M.; Levine, R.; Goodwin, A.; Pau
Submitted to the EMBL Data Library, April 1997
A:Description: Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial
A:Reference number: Z14590
A:Accession: T02205
A:Status: preliminary;
A:Molecule type: mRNA
A:Residues: 1-905<ELB>
A:Cross-references: EMBL:AF001261; NID:g2623762; PIDN:AAB86529.1; PID:g2623763

Tue Apr 2 09:40:10 2002

```
A:Experimental source: lung
A:Accession: T02152
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-793 'SGSFP', 'M', 862, 'RF', 865-866, 'Q', 868, 'AKVLELO', 876, 'QHQ', 880, 'FQ', <EL2>
A:Cross-references: EMBL:AF001263; NID:g2623766; PIDN:AAB86531.1; PID:g2623767
A:Experimental source: lung
A:Accession: T02171
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-792 'ES', <EL3>
A:Cross-references: EMBL:AF001262; NID:g2623764; PIDN:AAB86530.1; PID:g2623765
A:Experimental source: lung

alignment_scores:
  Quality: 233.00      Length: 83
  Ratio: 3.236        Gaps: 0
  Percent Similarity: 86.747      Percent Identity: 54.217

alignment_block:
US-09-049-696-6 x T02205
  Align seg 1/1 to: T02205 from: 1 to: 905

  1 CAAGAATTCTGTTCCTTGTAGTCTTGCACAAATCTGGAAGCATGGCGACTGG 50
  307 GlnArgValValCysLeuValLeuAspLysSerGlySerMetSerAlaG1 323
  51 TAACCGCCTCAATCGACTGAATCAAGCAGCCAGCTTTCTCTGCTGCAGA 100
  323 uAspArgLeuPheGlnMetAsnGlnAlaAlaGluLeuTyrLeuIleGlnV 340
  101 CAGTTGAGCTGGGTCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGC 150
  340 allLeGluLysGlySerLeuValGlyMetValThrPheAspSerValAla 356
  151 CATGTACAAAGTCAACATACACATAAAGTGGCAGTGGCAGTGACAGGACAC 200
  357 GluileGlnAsnHisLeuThrArgIleThrAspAsnValTyrGlnLy 373
  201 ACTGCCAAAAGATTACTCGCAGCAGCTTCAGGAGGACGTCCTCATCGC 249
  373 sileThrAlaLysLeuProGlnValAlaAsnGlyGlyThrSerIleCys 389

seq_name: pir2:S76691

seq_documentation_block:
hypothetical protein - Synecocystis sp. (strain PCC 6803)
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S76691
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S76691
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-420 <KAN>
A:Cross-references: EMBL:D64004; GB:AB001339; NID:g1001701; PIDN:BAAL0635.1; PID:di01128
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

alignment_scores:
  Quality: 87.00      Length: 79
  Ratio: 1.673        Gaps: 3
  Percent Similarity: 65.823      Percent Identity: 35.443

alignment_block:
US-09-049-696-6 x S76691
  Align seg 1/1 to: S76691 from: 1 to: 420

  10 GTGTGTTTGTCTTGTGCAATCTGGGAAGCATGGCGACTGTAACCGCCT 59
  44 LeuCysLeuValLeuAspHisSerGlySerMet...AspGlyGlnProle 59
  60' CAATCGACTGAATCAAGCAGCCAGCTTTCTCTGCTGCAGACAGATTGAGC 109
  59 uLThrValLysSerAlaAla...LeuGlyLeuIleAspArgLeuGluG 75
  110 TGGGGTCTGGGTGGGATGTCAGATTTGACAGTGTGCTGCCCATGTACAA 159
  75' luAspAspArgLeuSerValIleAlaPheAspHisArgAlaLysIleVal 91
  160 AGTGAACATACAGATAAAGTGGCAGTGACAGGGACACACTCGCCAA 209
  92 IleGluAsnGlnGlnValArgAsnGlyAlaAlaIleAlaLysAlaIleG1 108
  210 AAGATTACCTGCAGCAGCTTCAGGAGGAGCAGTCCATC 246
  108 uArgLeu.....LysAlaGluGlyGlyThrAlaIle 118

seq_name: pir2:H81800

seq_documentation_block:
conserved hypothetical integral membrane protein NMA1760 [imported] - Neisseria mening
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: H81800
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491
A:Reference number: A81775; MUID:20222556
A:Accession: H81800
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-356 <PAR>
A:Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB84988.1; PID:g738
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: NMA1760

alignment_scores:
  Quality: 74.00      Length: 60
  Ratio: 1.762        Gaps: 1
  Percent Similarity: 70.000      Percent Identity: 31.667

alignment_block:
US-09-049-696-6 x H81800
  Align seg 1/1 to: H81800 from: 1 to: 356

  16 TTAGTCCTTGACAAATCTGGAAGCATGGCGACTGTTAACCCTCAATCG 65
  86 LeuThrValIleLysAlaSerGlyMetSerThrLysLysLeuLeuIle 102
  66 ACTGAATCAAGCAGCCAGCTTTCTCTGCTGCAGACAGATTGAGCTGGGT 115
  102 eLeuSerGlnPheGlyPheIleAlaIleAlaThrValAlaLeuGlyG 119
  116 CTGTGGTGGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 165
  119 lutrPValAlaProThrLeuSerGlnLysAlaGluAsnIleLysAlaAla 135
  166 CTCATA.....CAGATAAACAGTGGCAGT 189
  136 AlaIleAsnGlyLysIleSerThrGlyAsn 145

seq_name: pir2:F81068
```

seq_documentation_block:
 C:Species: Neisseria meningitidis
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C:Accession: F81068
 R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
 ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
 Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V
 A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
 A:Reference number: AB1000; MUID:20175755
 A:Accession: F81068
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-356 <FET>
 A:Cross-references: GB:AE002507; GB:AE002098; NID:g7226811; PIDN:AAF41924.1; PID:g722681
 A:Experimental source: serogroup B, strain MC58
 C:Genetics:
 A:Gene: NMB1571

alignment_scores:		
Quality:	74.00	Length: 60
Ratio:	1.762	Gaps: 1
Percent Similarity:	70.000	Percent Identity: 31.667

alignment block:

US-09-049-696-6 x F81068

Align seq 1/1 to: F81068 from: 1 to: 356

16 TTAGTCCTTGACAAATCTGGAGAGATGGCGACTGGTAACCGGCTCAATCG 65
 |||||::: |||||:::|||||:::|||||::: |||||
 86 LeuThrValIIeLysAlaSerGlyMetSerThrLysLysLeuLeuII 102
 |||||:::|||||:::|||||:::|||||:::|||||
 66 ACTGAATCAACGAGCCACGCTTTTCCTGCTGCAGACAGCTTGAGCTGGGT 115
 |||||:::|||||:::|||||:::|||||:::|||||
 102 eLeuSerGlnPheGlyPheIIePheAlaIIeAlaThrValAlaLeuGlyG 119
 |||||:::|||||:::|||||:::|||||:::|||||
 116 CTTGGTTTGGGATGGTGACATTTCCACAGTCTGCCCATGTACAAAGCTGAA 165
 |||||:::|||||:::|||||:::|||||:::|||||
 119 IuTrpValAlaProThrLeuSerGlnLysAlaGluAsnIIeLysAlaAla 135
 |||||:::|||||:::|||||:::|||||:::|||||
 166 CTCATA.....CAGATAAACAGTGGCAGT 189
 |||||:::|||||:::|||||:::|||||:::|||||
 136 AlaIIeAsnGlnLysIIeSerThrGlyAsn 145

seq name: pir2:D82493

seq_documentation_block:
conserved hypothetical protein VCA0171 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)
C:Species: *Vibrio cholerae*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: D82493
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Church, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.; White, O.; Salzberg, S.L.; Peterson, J.D.; Fouts, D.E.; Karp, P.; Klenk, A.P.; Koc, A.; Kuskopfs, B.; Lior, H.; Olson, R.G.; Peterson, J.D.; Ravel, R.R.; Mekalanos, J.D.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A:Reference number: A82035; MUID:20406833
A:Accession: D82493

A:Status: preliminary

A; Molecule type: DNA

A;Residues: 1-646 <HEI>

A; Cross-references: GB:

A; Experimental source:

C;Genetics:

A:Gene: VCA0171
A:Map position: 3

A; Map position: 2

```

alignment_scores:      71.50      Length:      71
                       Quality:    1.702      Gaps:      4
                       Ratio:      59.155     Percent Identity: 42.254

alignment_block:
US-09-049-696-6 x D82493  ..

Align seg 1/1 to: D82493 from: 1 to: 646

      16  TTAGTCTTGACAAATCTGAAGCATG...GCGACTGGTAACCGCCTCAA  62
      80  LeuilewecAspMetserArgSerMetYrAlatirAspLeuAlaProAs  105
      69  TCGACTGAATCAAGCA.....GCCACGCTTTTCTGCTGCAGACAGATTG  106
      105  nArgLeuThrGlnAlaArgYrLysAlaLeuAspLeuLeuLysGlyTrp  122
      107  AGCTGGGGTCTCGGGTTGGGATGGTGCACATTGACAGTCGTCGCCCATGTA  156
      122  lnglucIvser...ThrGlyLeuValAlaTrAlaAlaAsoAlaTrVal  137

```

[illegible]

149	aAsnI.euI.euPro	153
-----	-----------------	-----

CT 01272115V 641

```
seq_name: p1r2:JC5953
```

```

seq_documentation_block:
  inter-alpha-inhibitor H4P heavy chain - rat
  C:Species: Rattus norvegicus (Norway rat)
  C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999
  C:Accession: JC5953
  N:Soury, E.; Olivier, E.; Daveau, M.; Hiron
  Biochem. Biophys. Res. Commun. 243, 522-530
  A:Title: The H4P heavy chain of inter-alpha
  A:Reference number: JC5953; MUID:98153798
  A:Accession: JC5953
  N>Status: preliminary
  A:Molecule type: mRNA
  A:Residues: 1-932 <SD>
  A:Cross-references: GB:Y11283; NID:G2292987
  C:Superfamily: inter-alpha-trypsin inhibitor

```

alignment scores:

```

      11  Quality: 70.50      Length: 88
      12  Ratio: 1.195      Gaps: 6
      13  Percent Similarity: 67.045      Percent Identity: 29.545

alignment_block:
US-09-049-696-6 x JC5953      ..

Align_seg 1/1 to: JC5953 from: 1 to: 932

      1  AAGAAATGTGTGTTTACTCTGCACAAATCTGGAAGCATGCGACTGGTA 52
      2  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
      3  :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
      4  272 LysAsnValLeuPhe.ValIleAspLysSerGlySerMeta...GlyL 287

```

53 ACCEPTED

ACCGCTC

287 yslYsile
.....

11-11-61

103 GTGAGCT

• • • • •

303 LeuSerTh

SECRET

158 TGTACAAA

11

100

21

—


```
A:Molecule type: DNA
A:Residues: 1-452 <DUZ>
A:Cross-references: EMBL:U53180; PIDN:AAA96290.1; GSPDB:GN00023; CESP:D1014.1
A:Experimental source: strain Bristol N2; clone D1014
C:Genetics:
A:Gene: CESP:D1014.1
A:Map position: 5
A:Introns: 256/3; 336/2; 376/3; 405/3

alignment_scores:
  Quality: 67.00      Length: 42
  Ratio: 2.310        Gaps: 0
Percent Similarity: 69.048 Percent Identity: 35.714

alignment_block:
US-09-049-696-6/rev x T29618 ..
Align seg 1/1 to: T29618 from: 1 to: 452

128 ATCCACACCCAGACCCAGCTCAACTGTGTGAGCAGGAGAAAGCTGGCC 79
|||||
407 ILleProCysAsnProProSerCysLysCysAspLysLeuSerArgPr 423
|||||
78 TGCTTGATTCAGTCGATTGAGCGGTTACCAGTCGCCCATGCTTCCAGATT 29
|||||
423 OThrGluPheAspGluSerArgProAspTyrValGlyLeuValProAspL 440
|||||
28 TGTCAGGACTAAACACAAATCTT 3
|||||
440 euGluLysThrGluTyrGluLeuLeu 448

seq_name: pir2:F72453

seq_documentation_block:
probable cytochrome C-type biogenesis protein APE2274 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: F72453
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; KDNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum
A:Reference number: A72450; MUID:99310339
A:Accession: F72453
A:Molecule type: DNA
A:Residues: 1-999 <KAW>
A:Cross-references: DDBJ:AP000064; NID:g5105945; PIDN:BAA81286.1; PID:d1045072; PID:g510
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE2274

alignment_scores:
  Quality: 67.00      Length: 49
  Ratio: 2.233        Gaps: 1
Percent Similarity: 61.224 Percent Identity: 40.816

alignment_block:
US-09-049-696-6/rev x F72453 ..
Align seg 1/1 to: F72453 from: 1 to: 999

236 CTCCTGAGCTGTCGAGGTAACTT.....TTGGC 205
|||||
325 ProProArgIleAlaGluMetAsnLeuSerAsnIleLeuLeuThrValGI 341
|||||
204 GAGTGTGTCCTGTCAGTCGCCACTCTTTATCTGTATGATGAGTTCACTTTGT 155
|||||
341 ySerValSerIleAlaLeuSerAlaLeuIleValTyrSerSerLeuLeuT 358
|||||
154 CATGGGAGCACTGTCAAATGTCAACATCCCAACCCAGGCCAGC 108

|| .....|||: ||| :||| |||||
358 hrProSerValLeuLysLeuValGlyLeuAspThrGlnProSer 373

seq_name: pir2:E70589

seq_documentation_block:
protoporphyrin IX magnesium chelatase (EC 4.99.1.-) - Mycobacterium tuberculosis (str
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: E70589
R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Keltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: E70589
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-629 <COL>
A:Cross-references: GB:295207; GB:AL123456; NID:g3261745; PIDN:CAB08456.1; PID:e31518
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV2850c
C:Keywords: lyase

alignment_scores:
  Quality: 66.50      Length: 84
  Ratio: 1.415        Gaps: 1
Percent Similarity: 55.952 Percent Identity: 27.381

alignment_block:
US-09-049-696-6 x E70589 ..
Align seg 1/1 to: E70589 from: 1 to: 629

7 ATGTGTGTTAGTCTTGACAAATCTGAAGCATGGCGACTGTTAACCG 56
|||||
451 LeuValIlePheValValAspAlaSerGlySerMetAlaAlaArgAspAr 467
|||||
57 CCTCAATCGACTCAATCAAGCAGCGCCAGCTTTCTGCTGCGACAGAGTTG 106
|||||
467 gMetAlaAlaValSerGlyAlaThrLeuSerLeuLeuArgAspAlaTyrG 484
|||||
107 AGCTGGGTCCTGGTGGGATGGTGACATTT..... 138
|||||
484 InArgArgAspLysValAlaValIleThrPheArgGlnHisGluAlaThr 500
|||||
139 .....GACAGTGTGCCCATGTACAAAGTGAACACTCATACA 173
|||||
501 LeuLeuLeuSerProThrSerSerAlaHisIleAlaGlyArgArgLeuAl 517
|||||
174 GATAAACAGTGGCAGTCACAGGACACACTCGCCCAAAAGATTACCTGCAG 223
|||||
517 aArgPheSerThrGlyGlyLysThrProLeuAlaGluGlyLeuLeuAlaA 534
|||||
224 CA 225
534 la 534
```

OM of: US-09-049-696-6 to: Issued_Patents_AA:* out_format : pfs
Date: Mar 30, 2002 2:32 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=framet_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09049696/runat_28032002_145238_2012/app_query.fasta_1.112579
-DB=Issued_Patents_AA -QFMT=Fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696@cgn1_1_109 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-6
Query length: 252
Database: Issued_Patents_AA:*
Database sequences: 212252
Database length: 22503292
Search time (sec): 362.110000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-09-188-930-305			70.00	138.96	649
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-447-411-76			66.50	123.00	1333
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-662-227-34			66.50	123.00	1333
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-017-947-34			66.50	123.00	1333
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-804-227C-5			66.50	121.04	1611
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-804-227C-4			66.50	112.37	3729
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-804-227C-2			66.50	110.50	2.37
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-944-483-47			63.00	131.73	2.77
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-980-115-14			63.00	118.16	4.25
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-977-554-2			62.50	125.88	3.75
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-225-867-2			62.50	125.88	3.75
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-227-806-2			62.50	125.88	3.75
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-985-526-1			62.00	130.75	3.62
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-985-526-1			62.00	129.80	3.73
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-09-111-556A-5			62.00	127.55	4.00
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-360-738-5			62.00	127.55	4.00
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-360-738-5			62.00	124.69	4.38
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-313-288B-20			61.50	113.39	6.25
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-483-857-8			61.50	127.51	4.51
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-861-774E-90			61.50	123.36	5.13
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-063-552-13			61.50	120.67	5.59
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-063-552-13			61.50	120.67	5.59
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-063-552-4			61.50	120.65	5.59
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-766-982-11			59.50	124.20	8.01
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-944-483-55			59.50	124.20	8.01
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-09-296-219-11			59.50	124.20	8.01
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-184-012C-8			59.50	112.45	11.60
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-334-177-2			59.50	112.45	11.60
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-666-082B-1			59.50	112.45	11.60
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-766-982-2			59.50	112.45	11.60
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-296-219-2			59.50	112.45	11.60
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-985-526-3			59.00	116.16	11.61
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-132-990A-8			59.00	112.50	13.03
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-477-451-3			59.00	95.77	22.05
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-415-818-6			58.50	116.73	12.83
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-894-236-6			58.50	116.73	12.83

/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-919-624-4 - 58.50 116.73 12.83 371
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-919-624-4 - 58.50 116.73 12.83 371
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-164-292B-26 + 58.00 105.45 20.59 983
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-845-623-26 + 58.00 105.45 20.59 983
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-845-623-26 + 58.00 105.45 20.59 983
seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-188-930-305
seq_documentation_block:
; Sequence 305, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 305
; LENGTH: 649
; TYPE: PRT
; ORGANISM: Mouse
US-09-188-930-305

alignment_scores:

Quality	70.00	Length	73
Ratio	1.750	Gaps	3
Percent Similarity	54.795	Percent Identity	30.137
alignment_block:			
US-09-049-696-6/rev x US-09-188-930-305			
Align seg 1/1 to: US-09-188-930-305 from: 1 to: 649			
233 CCTCAGTCGTCGAGTAACTTTTGGCGAGTGTGTCCTGCTGCTGCC 184			
337 ProglulysValArgGly.....MetAlaIleLysAspLeuSerAlaGl 351			
183 ACTGTTTATCTGTATGAGTTCACCTTTGTACATGGCGACGACTGTCAAATG 134			
351 uLeuPheAspCysLysAspSer.....GlyIleValSerThr 364			
133 TCACATCCCAACCCAGCAGCCAGCTCACTGCTGCAGCAGGAGAAAGC 84			
361 leGlnIleThrThrAlaIleProAsnThrAlaTyProAlaGlnGlyGln 380			
83 TGGCTGCTGCTGATTGATTCAGTCGAGCGGTACCGTCGCGCATGCTTCC 34			
361 TrpProAla.....ProValThrLysGlnPr 389			
33 AGATTTCACAGGACTAAA 15			
389 oAspIleLysAsnProLys 395			
seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-447-411-76			
seq_documentation_block:			
; Sequence 76, Application US/08447411			
; Patent No. 5773243			
; GENERAL INFORMATION:			
; APPLICANT: FRITZINGER, DAVID C.			
; APPLICANT: BREDEHORST, REINHARD			
; APPLICANT: VOGEL, CARL-WILHELM			
; TITLE OF INVENTION: DNA ENCODING COBRA C3, CVF1, AND CVF2			
; NUMBER OF SEQUENCES: 81			
; CORRESPONDENCE ADDRESS:			

NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCELLELAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/662,227
FILING DATE: 14-JUN-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 1126-0107-0X
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 1333 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-662-227-34

ADDRESSEE: OBLON, SPIVAK, MCCELLELAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,411
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/043,747
FILING DATE: 07-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5773243man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 1126-101-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
INFORMATION FOR SEQ ID NO: 76:
SEQUENCE CHARACTERISTICS:
LENGTH: 1333 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-447-411-76

alignment_scores:
Quality: 66.50 Length: 54
Ratio: 1.797 Gaps: 1
Percent Similarity: 68.519 Percent Identity: 27.778

alignment_scores:
Quality: 66.50 Length: 54
Ratio: 1.797 Gaps: 1
Percent Similarity: 68.519 Percent Identity: 27.778

alignment_block:
US-09-049-696-6 x US-08-662-227-34 ..
Align seg 1/1 to: US-08-662-227-34 from: 1 to: 1333
52 AACCGCCTCAATCGACTGAATCAAGCAGCCAGCTTTTCTGCTGCAGAC 101
234 SerArgAspAsnArgileGlnMetProGlyAlaAlaMetLysileLysLe 250
102 AGTT...GAGCTGGGTCCTGGGTTGGGATGGTGACATTTGACAGTGCTG 148
250 uGluglyAspProGlyAlaTrpIleGlyLeuValAlaValAspLysAlaG 267
149 CCCATGTACAAAGTCAACTCATACAGATAAACAGTGGCAGTCACAGGAC 198
267 lutyValLeuAsnAspLysTyrlsileSerGlnAlaLysileTrpAsp 283
199 ACACTCGCCAAA 210
284 ThrileGluLys 287

alignment_block:
US-09-049-696-6 x US-08-447-411-76 ..
Align seg 1/1 to: US-08-447-411-76 from: 1 to: 1333
52 AACCGCCTCAATCGACTGAATCAAGCAGCCAGCTTTTCTGCTGCAGAC 101
234 SerArgAspAsnArgileGlnMetProGlyAlaAlaMetLysileLysLe 250
102 AGTT...GAGCTGGGTCCTGGGTTGGGATGGTGACATTTGACAGTGCTG 148
250 uGluglyAspProGlyAlaTrpIleGlyLeuValAlaValAspLysAlaG 267
149 CCCATGTACAAAGTCAACTCATACAGATAAACAGTGGCAGTCACAGGAC 198
267 lutyValLeuAsnAspLysTyrlsileSerGlnAlaLysileTrpAsp 283
199 ACACTCGCCAAA 210
284 ThrileGluLys 287

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-017-947-34
seq_documentation_block:
; Sequence 34, Application US/09017947
; Patent No. 6303754
; GENERAL INFORMATION:
; APPLICANT: VOGEL, CARL-WILHELM
; APPLICANT: BREDEHORST, REINHORST
; APPLICANT: KOCK, MICHAEL
; APPLICANT: FRITZINGER, DAVID
; TITLE OF INVENTION: RECOMBINANT PROCVF
; NUMBER OF SEQUENCES: 39

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-662-227-34
seq_documentation_block:
; Sequence 34, Application US/08662227
; Patent No. 5922320
; GENERAL INFORMATION:
; APPLICANT: VOGEL, CARL-WILHELM
; APPLICANT: BREDEHORST, REINHORST
; APPLICANT: KOCK, MICHAEL
; APPLICANT: FRITZINGER, DAVID
; TITLE OF INVENTION: RECOMBINANT PROCVF

```

TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1611 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-804-227C-5

```

```
align ment scores:
+ + + + +
Quality: 66.50 Length: 63
Ratio: 1.900 Gaps: 2
Percent Similarity: 55.556 Percent Identity: 31.746
```

Percent similarity: 33.536 Percent identity: 31.746
alignment block:
US-09-p49-696-6 x US-08-804-227C-5 ..

Align:seg 1/1 to: US-08-804-22/C-5 from: 1 to: 1611

44	CGACTGTAACCGCCCTCAATCGACTCAATCAACGCGCAGCTTTTCCTG	93
884	AradLeuThrSerThrAlaGln	AlaThrAlaHis

94 CTGCAGACATTGACGTGGGGCTCCTCGGTTGGATCGTCACATTTCACAG 143
|||||
895 sGIYAlaThrLeuThrTrpAspProAlaLeuProGlyHisLeuThrT 912

144 TGCTGCCATGTACAAAGTCACTACAGA..... 175
 :||||||| ||| :|:|:|:
912 hrLeuProThrTyrProPheAsnHisHisTyrTrpLeuAspThrThr 928

176 TARACACTGCCACTGACAGGACACACTGCCCAAAAGAT 214
||||: ||||: ||||: ||||: ||||: ||||
929 proThrThrProalathrThrGlnSerProThrAsp 941

```
seq_name: /cgn2_b/ptodata/2/1aa/5B_COMB.pep:US-08-804-227C-4
      :
      :
seq_documentation_block:
      : Sequence 4, Application US/08804227C
      : Patent No. 5876991
```

GENERAL INFORMATION:
 APPLICANT: DeHoff, Bradley S.
 APPLICANT: Kuhstoss, Stuart A.
 APPLICANT: Rostock, Paul R., Jr.

APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:

ADDRESSEE: THOMAS G. PLANT 1501

alignment_scores:		
Quality:	66.50	Length: 63
Ratio:	1.900	Gaps: 2
Percent Similarity:	55.556	Percent Identity: 31.746

```

alignment_block:
US-09-049-696-6 x US-08-804-227C-2      ..
Align seg 1/1 to: US-08-804-227C-2 from: 1 to: 4472

```

```

44  CGACTGGTAAACGGCTCAATCGACTCAATCAAGCAGCGCACTTTTCCCTG 93
      |||||:||||:||||: |||: |||||: |||||: |||||: |||||:
1903 ArgLeuLeuThrSerThrAlaGlu.....AlaTrpAlaHis 1919

      94  CTGCAGACAGTTGAGCTGGGGTCTCGGTGGATGGTGACATTTGACAG 143
      |||||:||||:||||: |||||: |||||: |||||: |||||: |||||:
1919 sGlyAlaThrLeuThrTrpAspProAlaLeuProGlyHisLeuThr 1936

      144. TCGTCGCCATGTACAAAGTGAACATCATACAGA..... 175
      |||||:||||| ||| |||||: |||||: |||||: |||||:
1936 hrLeuProThrTyrProPheAsnHisHisTyrTrpLeuAspThrThr 1952

      176  TAAACAGTGGCAGTGCACGGACACACTGCCCAAAAGAT 214
      ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
1953 ProThrThr-ProAlaThrThrThrGlnSerProThrAsp 1965

seq name: /cgn2_6/otodata/2/1aa/6B_COMB_dep:US-08-944-483-47

```

```

1 seq_documentation_block:
2 ; Sequence 47, Application US/089444483
3 ; Patent No. 6232456
4 ;
5 ; GENERAL INFORMATION:
6 ;
7 ; APPLICANT: COHEN, MAURICE
8 ; APPLICANT: COLPITTS, TRACEY L.
9 ; APPLICANT: FRIEDMAN, PAULA N.
10 ; APPLICANT: GRANADOS, EDWARD N.
11 ; APPLICANT: KLASS, MICHAEL R.
12 ; APPLICANT: RUSSELL, JOHN C.
13 ; APPLICANT: STEWART, KENT D.
14 ; APPLICANT: STROUPE, STEVEN D.
15 ; TITLE OF INVENTION: NOVEL SERINERIN
16 ; TITLE OF INVENTION: AND METHODS
17 ; TITLE OF INVENTION: OF THE PROST
18 ;
19 ; NUMBER OF SEQUENCES: 76
20 ;
21 ; CORRESPONDENCE ADDRESS:
22 ; ADDRESSEE: Abbott Laboratories
23 ; STREET: 100 Abbott Park Road
24 ; CITY: Abbott Park

```

APPLICANT: STROUPE, STEVEN D.
TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
TITLE OF INVENTION: OF THE PROSTATE

```
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 251 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
US-08-944-483-47

alignment_scores:
  Quality: 63.00      Length: 69
  Ratio: 1.909       Gaps: 4
Percent Similarity: 47.826 Percent Identity: 33.333

alignment_block:
US-09-049-696-6 x US-08-944-483-47

Align seg 1/1 to: US-08-944-483-47 from: 1 to: 251

25 GACAAATCTGGAGCAGTGGCGACTGGTAACGCCTCAATGACTGAATCA 74
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
197 AspSerGlyGlyPrometValThrLeuAsnArg.....G1 208

75 AGCAGGCGCAGCTTTCTGCTGCAGACAGTTGAGCTGGGTCTCTGGGTTG 124
: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
208 uArgGlyGlnTrpTyrLeuValGlyThrVal.SerIrpGlyAspAspCys 224

125 GGATGGTGACATTGTGACAGTGTCTGCCCATGTACAAAGTGAACATACAG 174
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
225 Gly.....LysLysAspArgTyrGlyValTyrSerTyr... 235

175 ATAAACAGTGGCAGTGACAGGGACACACTGCCCAAAAGATTACCTGCAGC 224
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 .....IleHisHisAsnLysAspTrpIleGlnA 245

225 AGCTT 229
: |||:
245 rgVal 246

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-764-870-14

seq_documentation_block:
; Sequence 14, Application US/08764870
; Patent No. 6236946
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Thomas S
; APPLICANT: Baxter, John D
; APPLICANT: Fletcher, Robert J
; APPLICANT: Wagner, Richard L

; APPLICANT: Kushner, Peter J
; APPLICANT: Apriletti, James W
; APPLICANT: West, Brian
; TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward
; STREET: Five Palo Alto Square, 3000 El Camino Real
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,870
; FILING DATE: 13-DEC-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,540
; FILING DATE: 13-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,543
; FILING DATE: 13-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,606
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakamura, Jackie N
; REGISTRATION NUMBER: 35,966
; REFERENCE/DOCKET NUMBER: UCAL-246/01US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)843-5000
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 933 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-764-870-14

alignment_scores:
  Quality: 63.00      Length: 88
  Ratio: 1.286       Gaps: 4
Percent Similarity: 55.682 Percent Identity: 32.955

alignment_block:
US-09-049-696-6 x US-08-764-870-14

Align seg 1/1 to: US-08-764-870-14 from: 1 to: 933

13 TGTTCCTTCCTGACAAATCTGGAAGCATGCGACTGGTAACCGCTCAA 62
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
170 CysLysValGlyAspSerSerGlyThrAlaAlaHisLysValLeuPr 186

63 TCGA...CTGAATCAACAGCCAGCTTTCTCTGCTGCAGAGTTGAGC 109
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 oArgGlyLeuSerProAlaArgGlnLeu...LeuLeuProAlaSerGlu 202

110 TGGGGTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 159
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
202 erProHisTrpSerGlyAlaProValLysProSerProGlnAlaAla 218

160 AGTGAACATCATACAGATAACAGTGGCGAGTGCAGGGGACACA..... 201
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
210 ValGluValGluGluGluAspSerSerGluSerGluSerAlaGlyPr 235
```

```

202 .CTCGCAAAAGATTACCT.....GCAGCAGCTTCAGGAG 235
      ||| ||| |||
235 oLeuLeuLysGlyLysProArgAlaLeuGlyGlyAlaAlaGlyGlyC 252
      ||| ||| |||
236 GGAGCTGCATCTGCC 249
      ||:|:|:| |||
252 lValAlaAlaCys 256

```

seq name: /cqn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-980-115-14

```

seq_documentation_block:
; Sequence 14, Application US/08980115
; Patent No. 6266622
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Thomas S.
; APPLICANT: Baxter, John D.
; APPLICANT: Fletterick, Robert J.
; APPLICANT: Wagner, Richard L.
; APPLICANT: Kushner, Peter J.
; APPLICANT: Apriletti, James W.
; APPLICANT: West, Brian L.
; APPLICANT: Shlau, Andrew K.
; TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS
; FILE REFERENCE: UCAL-245/020US
; CURRENT APPLICATION NUMBER: US/08/980,115
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: 08/764,870
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 60/008,606
; EARLIER FILING DATE: 1995-12-14
; EARLIER APPLICATION NUMBER: 60/008,543
; EARLIER FILING DATE: 1995-12-13
; EARLIER APPLICATION NUMBER: 60/008,540
; EARLIER FILING DATE: 1995-12-13
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 933
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (659)..(918)
; OTHER INFORMATION: minimal ligand binding domain
; US-08-980-115-14

```

```

Alignment_block:
US-09-049-696-6 x US-08-980-115-14 ..
Align seq 1/1 to: US-08-980-115-14 from: 1 to: 933

```

```

alignment_scores:      Quality:      62.50      Length:      98
                       Ratio:        1.136      Gaps:         3
                       Percent Similarity: 56.122      Percent Identity: 22.449

alignment_block:
.US-09-049-696-6 x US-08-977-554-2 ..
Align seg 1/1 to: US-08-977-554-2 from: 1 to: 393

..7 ATTCTGCTGTTTACTCTCTTGACAAATCTGGAAACGATGGCGACTGGTAACCG 56

```



```
160 ValIleCysGluIleMetAsnAspGlyThrMetAlaLysGlyGlnAs 176
57 CCTCAATCGACTGAATCAAGCAGCCAGCTTTTCTCTGCTG..... 96
176 pLeuGlnAsnPhelYsGluLysHisGlnLeuLysMetIleThrIleAspA 193
97 .....CAGACAGTTGAGCTGGGTGCC 117
193 sPLeuIleGluTyArgLysLysLeuGluProGluIleGluPheLysAla 209
118 TGGTGTGGATGGTG.....ACATTGTGAC.....AG 143
210 LysValLysMetProThrAspPheGlyThrPheAspMetTyrglyPheLy 226
144 TGCTGCCCATGTACAAGTGAACATACAGATAAACAGTGGCAGTGACA 193
226 sAlaThrTyThrAspGluGluIleValLeuThrLysGlyAlaIleA 243
194 GGGACACACTCGCCAAAGATTACCTGCAGCAGCTTCAGGAGGG 237
243 rGlnHisGluAsnValArgLeuHisSerAlaCysLeuThrGly 257
seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-225-967-2
```

```
seq documentation block:
; Sequence 2, Application US/09225967
; Patent No. 6171598
; GENERAL INFORMATION:
; APPLICANT: Palmer, Leslie M.
; APPLICANT: Fedon, Jason C.
; APPLICANT: Warren, Richard L.
; APPLICANT: Trainl, Christopher M.
; APPLICANT: Wang, Min
; APPLICANT: Jaworski, Deborah D.
; APPLICANT: Mooney, Jeffrey
; APPLICANT: Debouck, Christine
; APPLICANT: Zhong, Yiyi
; APPLICANT: Black, Michael
; TITLE OF INVENTION: rIba
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/225,967
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/977,554
; FILING DATE:
; APPLICATION NUMBER: PCT/US97/02318
; FILING DATE: 19-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Q. Todd
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: P50444-07
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215/994-2252
; TELEFAX: 215/994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 393 amino acids
```

```
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-225-967-2
alignment_scores:
; Quality: 62.50 Length: 98
; Ratio: 1.136 Gaps: 3
; Percent Similarity: 56.122 Percent Identity: 22.449
alignment_block:
; US-09-049-696-6 x US-09-225-967-2
; Align seg 1/1 to: US-09-225-967-2 from: 1 to: 393
57 ATTGTGTTTGTAGCTTGCACAAATCTGGAAGCATGCGAGCTGGTAACCG 56
160 ValIleCysGluIleMetAsnAspGlyThrMetAlaLysGlyGlnAs 176
57 CCTCAATCGACTGAATCAAGCAGCCAGCTTTTCTCTGCTG..... 96
176 pLeuGlnAsnPhelYsGluLysHisGlnLeuLysMetIleThrIleAspA 193
97 .....CAGACAGTTGAGCTGGGTGCC 117
193 sPLeuIleGluTyArgLysLysLeuGluProGluIleGluPheLysAla 209
118 TGGTGTGGATGGTG.....ACATTGTGAC.....AG 143
210 LysValLysMetProThrAspPheGlyThrPheAspMetTyrglyPheLy 226
144 TGCTGCCCATGTACAAGTGAACATACAGATAAACAGTGGCAGTGACA 193
226 sAlaThrTyThrAspGluGluIleValLeuThrLysGlyAlaIleA 243
194 GGGACACACTCGCCAAAGATTACCTGCAGCAGCTTCAGGAGGG 237
243 rGlnHisGluAsnValArgLeuHisSerAlaCysLeuThrGly 257
seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-227-806-2
seq documentation block:
; Sequence 2, Application US/09227806
; Patent No. 6280971
; GENERAL INFORMATION:
; APPLICANT: Palmer, Leslie M.
; APPLICANT: Fedon, Jason C.
; APPLICANT: Warren, Richard L.
; APPLICANT: Trainl, Christopher M.
; APPLICANT: Wang, Min
; APPLICANT: Jaworski, Deborah D.
; APPLICANT: Mooney, Jeffrey
; APPLICANT: Debouck, Christine
; APPLICANT: Zhong, Yiyi
; APPLICANT: Black, Michael
; TITLE OF INVENTION: rIba
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/227,806
; FILING DATE:
```

```

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/977,554
FILING DATE:
APPLICATION NUMBER: PCT/US97/02318
FILING DATE: 19-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Q. Todd
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: P50444-07
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215/994-2252
TELEFAX: 215/994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 393 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-227-806-2

alignment_scores:
Quality: 62.50 Length: 98
Ratio: 1.136 Gaps: 3
Percent Similarity: 56.122 Percent Identity: 22.449

alignment_block:
US-09-049-696-6 x US-09-227-806-2

```

```

Align seg 1/1 to: US-09-227-806-2 from: 1 to: 393

7 ATTGTGTGTTTACTCCITGCAAAATCTGGAAGCATGGCGACTGTAACCG 56
  |||:||||| :|||:||||| :|||:||||| :|||:|||||
160 ValIecGlyGluIleMetAsnAspGlyThrMetAlaLysGlyGlnAs 176
  |||:||||| :|||:||||| :|||:||||| :|||:|||||
176 pLeuGlnAsnPhelYsGluLysHisGlnLeuLysMetIleThrIleAspA 193
  |||:||||| :|||:||||| :|||:||||| :|||:|||||
197 .....CAGACAGTTGAGCTGGGGTCC 117
  |||:||||| :|||:||||| :|||:||||| :|||:|||||
193 spLeuIleGluTyfArgLysLysLeuGluProGluIleGluPhelYsAla 209
  |||:||||| :|||:||||| :|||:||||| :|||:|||||
118 TGGGTTGGGATGGTG.....ACATTGAC.....AG 143
  |||:||||| :|||:||||| :|||:||||| :|||:|||||
210 LysValLysMetProThrAspPhelGlyThrPheAspMetTyfGlyPhelY 226
  |||:||||| :|||:||||| :|||:||||| :|||:|||||
144 TGCTGCCCATGTACAAAGTGAACTCATACAGATAACAGATGGCAGTGACA 193
  |||:||||| :|||:||||| :|||:||||| :|||:|||||
226 sAlaThrTyfThrAspGluIleValLeuValLeuThrLysGlyAlaIleA 243
  |||:||||| :|||:||||| :|||:||||| :|||:|||||
194 GGGACACACTCGCCAAAGATTACCTGCAGCAGCTTCAGGAGGG 237
  |||:||||| :|||:||||| :|||:||||| :|||:|||||
243 rgGlnHisGluAsnValArgLeuHisSerAlaCysLeuThrGly 257

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-985-526-1

seq_documentation_block:
; Sequence 1, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington

```

```

alignment_block:
US-09-049-696-6/rev x PCT-US93-01652-1      ..
Align seg 1/1   to: PCT-US93-01652-1 from: 1   to: 239

    208 TGGCGAGTGTGTCCTGT...CACTGCCACCATTT.....AT 175
        ||| ||| .....|||||||
    20  TtpThrValAspSerCysThrGlucyHisCysGlnAsnSerValThrIl 36

    174 CTCTATGAGTTCACITTTGTACATGGGCAGCACCTGTCAAATGTCACCATCC 125
        ||| ||| .....|||||||
    36  eCysLysLysValSerCysProIleMetProCysSerAsnAlaThrValP 53

    124 CAACCCAGGACCCCAGCTCAACTGTCTGCAGCAGGAAGAAGTGCGCTGCT 75
        ||| ..... |||
    53 rAspAsglyGlu.....CysCysProArgCysTrpProSer 64

    74 TGATTCACTCGATTGAGGC 56
        |||||.....|||
    65 AspSerAlaAspAspGly 70

```

This Page Blank (uspto)

Tue Apr 2 09:40:10 2002

us-09-049-696-6.rag

Align seg 1/1 to: AAG75614 from: 1 to: 869

```
1 CAAGAATTTCTGTTAGTCTTCAACAAATCTGAAGCATGGGACTGG 50
|||||
259 GlnArgIleValCysLeuValLeuAspLysSerGlySerMetAlaThrG1 275
51 TAACCGCTCAATCGACTGAATCAAGCAGCGGCGAGCTTTCTGCTGCAGA 100
|||||
275 yasnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheLeuGlnT 292
101 CAGTTGAGCTGGGCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 150
292 hrValGluLeuGlySerTrpValGlyMetValThrPheAspSerAla 308
151 CATGTACAAAGTCAACTCATACAGATAAAGTGGCAGTGACAGGACAC 200
|||||
309 HisValGlnSerGluLeuLeuGlnIleAsnSerGlySerAspArgAsp 325
201 ACTCGGCAAAAGATTACCTGCAGCAGCTTCAGGAGGAGCTCCATCTGCA 250
|||||
325 rLeuAlaLysArgLeuProAlaAlaSerGlyGlyThrSerIleCys 342
251 GC 252
342 er 342
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT: AAB74824

```
seq_documentation_block:
ID AAB74824 standard; Protein: 914 AA.
AC AAB74824;
DT 13-JUN-2001 (first entry)
DE Human ICACC-1 protein sequence.
KW ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW inflammatory bowel disease; autoimmune disease.
OS Homo sapiens.
PN WO9944620-A1.
PD 10-SEP-1999.
PF 03-MAR-1999; 99WO-US04703.
PR 03-MAR-1998; 98US-0076815.
PA (MAGA-) MAGAININ PHARM INC.
XX Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;
PI Nicolalides NC, Zhou Y, Dong Q;
XX WPI; 1999-550979/46.
DR N-PSDB; AAF81927.
XX New nucleic acid encoding calcium activated chloride channel, used to
PT identify, e.g. specific modulators for treating atopic allergy.
XX Claim 11; Fig 4B; 75pp; English.
XX The present sequence represents the human interleukin 9 (IL-9) induced
CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
CC have anti-allergic, anti-asthmatic, anti-inflammatory and
CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
CC used to alleviate asthma (or more generally atopic allergy), while those
CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
```

inflammatory bowel disease (IBD) (or other autoimmune diseases).
Measuring levels of ICACC-1 can be used in the diagnosis of asthma
(increased levels) or IBD (reduced levels), also for monitoring
treatment of these conditions. The ICACC proteins can be used:
(i) to raise specific antibodies (Ab), useful: (a) as immunoassay
reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
(A) to disrupt binding between ICACC and its ligands; and (iii) to
identify modulators and binding partners. ICACC polynucleotides can be
used to generate transgenic animals or recombinant cells, used to screen
for antagonists, also as a source of therapeutic antisense agents or
diagnostic probes (for quantifying mRNA expression, e.g. for
identification of modulators).

Sequence 914 AA:

```
alignment_scores:
Quality: 418.00 Length: 84
Ratio: 4.976 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-6 x AAB74824 ..
Align seg 1/1 to: AAB74824 from: 1 to: 914
1 CAAGAATTTCTGTTAGTCTTCAACAAATCTGAAGCATGGGACTGG 50
|||||
304 GlnArgIleValCysLeuValLeuAspLysSerGlySerMetAlaThrG1 320
51 TAACCGCTCAATCGACTGAATCAAGCAGCGGCGAGCTTTCTGCTGCAGA 100
|||||
320 yasnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheLeuGlnT 337
101 CAGTTGAGCTGGGCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 150
|||||
337 hrValGluLeuGlySerTrpValGlyMetValThrPheAspSerAla 353
151 CATGTACAAAGTCAACTCATACAGATAAAGTGGCAGTGACAGGACAC 200
|||||
353 HisValGlnSerGluLeuLeuGlnIleAsnSerGlySerAspArgAsp 370
201 ACTCGGCAAAAGATTACCTGCAGCAGCTTCAGGAGGAGCTCCATCTGCA 250
|||||
370 rLeuAlaLysArgLeuProAlaAlaSerGlyGlyThrSerIleCys 387
251 GC 252
387 er 387
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT: AAM24514

```
seq_documentation_block:
ID AAM24514 standard; Protein: 914 AA.
XX
AC AAM24514;
XX
DT 12-OCT-2001 (first entry)
XX C902P predicted amino acid sequence.
XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW immunogenic; gene therapy; vaccine; colonic cancer.
XX Homo sapiens.
OS
PN WO200149716-A2.
XX
PD 12-JUL-2001.
XX
PF 29-DEC-2000; 2000WO-US35596.
XX
XX 30-DEC-1999; 99US-0476296.
PR
```

```
PR 10-JAN-2000; 2000US-0480321.
PR 15-FEB-2000; 2000US-0504629.
PR 06-MAR-2000; 2000US-0519444.
PR 19-MAY-2000; 2000US-0575251.
PR 29-JUN-2000; 2000US-0609448.
PR 28-AUG-2000; 2000US-0649811.
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
XX WPI: 2001-441847/47.
XX
XX Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer -
PT
XX
XX Claim 2; Page 440-443; 472pp; English.
XX
XX The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be
CC used to treat disorders associated with decreased expression by
CC rectifying mutations or deletions in a patient's genome that affect the
CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host
CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512
CC and AA24494 to AA24523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.
XX
XX Sequence 914 AA;
SQ
alignment_scores:
Quality: 418.00 Length: 84
Ratio: 4.976 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-049-696-6 x AA24514
Align seg 1/1 to: AA24514 from: 1 to: 914
1 CAAGAATTGTGTTCCTTGCCTGACAAATCTGGAAGCATGGCGACTGG 50
|||||
304 GlnArgileValCysLeuValLeuAspLysSerGlySerMetAlaThrGl 320
51 TAACCGCTCAATCGACTGATCAAGCAGGCCAGCTTTCTCTGCTGCAGA 100
|||||
320 YAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheLeuLeuGlnT 337
101 CAGTTGAGCTGGGTCTCTGGTGGATGTGACATTTTGACAGTGCTGCC 150
|||||
337 hrValGluLeuGlySerTrpValGlyMetValThrPheAspSerAlaAla 353
151 CATGTACAAAGTCAACTCATACAGATAAAGCAGTGGCAGTGACAGGGACAC 200
|||||
354 HisValGlnSerGluLeuIleGlnIleAsnSerGlySerAspArgAspTh 370
201 ACTGCCAAAAGATTACCTGCAGCAGCTTCAGAGGAGGAGCTCCATCTGCA 250
|||||
370 rLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyThrSerIleCysS 387
251 GC 252
387 er 387
seq_name: /SDS2/gcdata/geneseq/geneseq/AA2001.DAT: AAB73716
seq_documentation_block:
ID AAB73716 standard; Protein; 914 AA.
XX
XX AC AAB73716;
XX
XX DT 11-SEP-2001 (first entry)
XX
XX DE Human CLCA1 protein, SEQ ID NO:2.
XX
XX KW Human CLCA1; goblet cell; mouse Gob-5 orthologue; drug screening;
KW expression inhibition; antisense therapy; gene therapy;
KW chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic.
XX
XX OS Homo sapiens;
XX
XX FH Key Location/Qualifiers
XX Key Location/Qualifiers
XX Misc-difference 152
XX
XX PN WO200138530-A1.
XX
XX PD 31-MAY-2001.
XX
XX PF 22-NOV-2000; 2000WO-JP08232.
XX
XX PR 24-NOV-1999; 99JP-0333479.
XX
XX PR 27-APR-2000; 2000JP-0127589.
XX
XX PA (TAKE ) TAKEDA CHEM IND LTD.
XX
XX PI Nakanishi A, Morita S;
XX
XX DR WPI: 2001-355935/37.
XX
XX DR N-SDB; AAH46102, AAH46124.
XX
XX New antisense nucleotide, useful for treatment and prevention of
XX bronchial asthma and chronic obstructive pulmonary disease .
XX
XX Claim 2; Page 76-80; 104pp; Japanese.
XX
XX The invention relates to an antisense nucleotide targetted to the mouse
XX Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
XX the CLCA1 gene (coding sequence shown in AAH46102). The invention also
XX relates to an antibody specific for the Gob-5 protein, medical and
XX diagnostic compositions containing the antisense nucleotide or the
XX antibody, and methods and kits for screening for compounds which inhibit
XX the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.
XX The antisense oligonucleotides and antibody are therefore useful for the
XX treatment and prevention of bronchial asthma and chronic obstructive
XX pulmonary disease. The present sequence represents human CLCA1 protein.
XX
XX Sequence 914 AA;
SQ
alignment_scores:
Quality: 418.00 Length: 84
Ratio: 4.976 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-049-696-6 x AAB73716
Align seg 1/1 to: AAB73716 from: 1 to: 914
```

CC antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;
CC vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic;
CC anticonvulsant; antialzheimer; antiparkinsonian; antimicrobial; and
CC vulnery Human secreted protein nucleotide sequences (NAM1) and proteins
CC vaccine. Human secreted protein nucleotide sequences (NAM1) and proteins
CC (PEP1) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. For example, NAM1
CC and PEP1 may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patients genome
CC that affect the activity of proteins by expressing inactive proteins or
CC to supplement the patients own production of polypeptides. Disorders that
CC may be prevented, diagnosed and/or treated include immune disorders,
CC hyperproliferative disorders (e.g. cancers), cardiovascular diseases,
CC angiogenic disorders, neurological disorders, infectious diseases and/or
CC for promoting wound healing, regeneration and/or chemotaxis. AAF81778 to
CC AAF81786 and AAB74732 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 914 AA;

alignment_scores:
Quality: 418.00 Length: 84
Ratio: 4.976 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-6 x AAB74733

Align seg 1/1 to: AAB74733 from: 1 to: 914

1 CAAGAATTGTGTTAGTCTTGACAAATCTGGAAGCATGGGACTGG 50
|||||
304 GlnArgIleValCysLeuValLeuAspLysSerGlySerMetAlaThrG1 320
51 TAACGCCCTCAATCGACTCAATCAAGCAGCGGCGCTTTCTGCTGCAGA 100
|||||
320 yAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheLeuLeuGlnT 337
101 CAGTTGAGCTGGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 150
|||||
337 hrValGluLeuGlySerTrpValGlyMetValThrPheAspSerAlaAla 353
151 CATGTACAAAGTGAACCTCATACAGATAACACAGTGGCAGTGACAGGACAC 200
|||||
354 HisValGlnSerGluLeuIleGlnIleAsnSerGlySerAspArgAspTh 370
201 ACTCGCCCAAAAGATTACCTGCGAGCAGCTTCAGGAGGAGCGTCCATCTGCA 250
|||||
370 rLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyThrSerIleCysS 387
251 GC 252
||
387 er 387

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAG75474

seq_documentation_block:
ID AAG75474 standard; Protein: 925 AA.

XX
AC AAG75474;

XX
DT 03-SEP-2001 (first entry)

XX
DE Human colon cancer antigen protein SEQ ID NO:6238.

XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 1.

XX
OS Homo sapiens.

XX
PN WO200122920-A2.

XX

1 CAAGAATTGTGTTAGTCTTGACAAATCTGGAAGCATGGGACTGG 50
|||||
304 GlnArgIleValCysLeuValLeuAspLysSerGlySerMetAlaThrG1 320
51 TAACGCCCTCAATCGACTCAATCAAGCAGCGGCGCTTTCTGCTGCAGA 100
|||||
320 yAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheLeuLeuGlnT 337
101 CAGTTGAGCTGGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 150
|||||
337 hrValGluLeuGlySerTrpValGlyMetValThrPheAspSerAlaAla 353
151 CATGTACAAAGTGAACCTCATACAGATAACACAGTGGCAGTGACAGGACAC 200
|||||
354 HisValGlnSerGluLeuIleGlnIleAsnSerGlySerAspArgAspTh 370
201 ACTCGCCCAAAAGATTACCTGCGAGCAGCTTCAGGAGGAGCGTCCATCTGCA 250
|||||
370 rLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyThrSerIleCysS 387
251 GC 252
||
387 er 387

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB74733

seq_documentation_block:

ID AAB74733 standard; Protein: 914 AA.

XX
AC AAB74733;

XX
DT 12-JUN-2001 (first entry)

XX
DE Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.

XX
KW Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
KW antialzheimer; antiparkinsonian; antimicrobial; vulnery; gene therapy;
KW immune disorder; hyperproliferative disorder; cardiovascular disease;
KW cancer; angiogenic disorder; neurological disorder; infectious disease;
KW wound healing; regeneration; chemotaxis; chromosome 1.

XX
OS Homo sapiens.

XX
PN WO200112775-A2.

XX
PD 22-FEB-2001.

XX
PF 16-AUG-2000; 2000WO-US22325.

XX
PR 17-AUG-1999; 99US-0149182.

XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX
PI Rosen CA, Ni J, Florence KA, Fiscella M, Wei P, Baker KP;
PI Birse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;

XX
DR WPI; 2001-147550/15.

XX
DR N-PSDB; AAF81787.

XX
PT Nucleic acids encoding 25 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -

XX
PS Claim 11; Page 459-460; 485pp; English.

XX
CC AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733
CC to AAB74772. Human secreted proteins can have activities based on the
CC tissues and cells they are expressed in. Example of activities include:
CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;


```
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI: 2001-235357/24.
DR N-PSDB; AAH34879.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
PT Claim 11; Page 7686-7690; 9803pp; English.
PS
XX AAH32943 to AAH37195 and AAG7788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 925 AA;

alignment_scores:
    Quality: 418.00      Length: 84
    Ratio: 4.976        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-049-696-6 x AAG75474

Align seg 1/1 to: AAG75474 from: 1 to: 925

1 CAAGAATGCTGTTAGTCCTTGACAAATCTGGAAGCATGCGACTGG 50
|||||
315 GlnArgileValcysLeuValLeuAspLysSerGlySerMetAlaThrG1 331

51 TAACCGCTCAATCGACTGAATCAAGCAGCCAGCTTTTCTCTGTCAGA 100
|||||
331 YAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheLeuLeuGlnT 348

101 CAGTTGAGCTGGGCTCCTCGTGGTGGATGCTGACATTGACAGTGCTGCC 150
|||||
348 hrValGluLeuGlySerTrpValGlyMetValThrPheAspSerAlaAla 364

151 CATGTACAAAGTGAACATCATACAGATAACACAGTGGCAGTGACAGGACAC 200
|||||
365 HisValGlnSerGluLeuIleGlnIleAsnSerGlySerAspArgAspTh 381

201 ACTCGCCAAAGATTACTCGACAGCTTCAGGAGGACGTCATCGCA 250
|||||
381 rLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyThrSerIleCysS 398

251 GC 252

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.AAB74822
seq_documentation_block:
ID AAB74822 standard; Protein; 913 AA.
XX
XX AC AAB74822;
XX
XX 13 JUN-2001 (first entry)
XX
XX Mifine ICACC-1 protein sequence.
XX
XX ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW inflammatory bowel disease; autoimmune disease.
XX
XX Mus sp.
XX
XX WO9944620-A1
XX
XX 10-SEP-1999.
PD
XX 03-MAR-1999; 99WO-US04703.
XX
XX 03-MAR-1999; 98US-0076815.
XX
XX (MAGA-) MAGAMIN PHARM INC.
XX
XX Hojroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;
PI Nicolaides NC, Zhou Y, Dong Q;
DR WPI: 1999-550979/46.
DR N-PSDB; AAF81925.
XX
XX New nucleic acid encoding calcium activated chloride channel, used to
PT identify, e.g. specific modulators for treating atopic allergy -
PS Claim 12; Fig 2; 75pp; English.
XX
XX The present sequence represents the murine interleukin 9 (IL-9) induced
CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
CC have anti-allergic, anti-asthmatic, anti-inflammatory and
CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
CC used to alleviate asthma (or more generally atopic allergy), while those
CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
CC (increased levels) or IBD (reduced levels), also for monitoring
CC treatment of these conditions. The ICACC proteins can be used:
CC (i) to raise specific antibodies (Ab), useful: (a) as immunosay
CC reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
CC identify modulators and binding partners. ICACC polynucleotides can be
CC used to generate transgenic animals or recombinant cells, used to screen
CC for antagonists, also as a source of therapeutic antisense agents or
CC diagnostic probes (for quantifying mRNA expression, e.g. for
CC identification of modulators).
XX
SQ Sequence 913 AA;

alignment_scores:
    Quality: 330.00      Length: 84
    Ratio: 4.231        Gaps: 0
    Percent Similarity: 92.857    Percent Identity: 77.381

alignment_block:
US-09-049-696-6 x AAB74822
```

Align seg 1/1 to: AAB74822 from: 1 to: 913

```
1 CAAAGAAATGTTGTTAGTCTTGACAAATCTGGAAGCATGGCGACTGG 50
|||||
305 GlnArgIleValCysLeuValLeuAspLysSerGlySerMetLeuAsnAs 321
51 TAACCGCCTCAATCGACTGAATCAAGCAGCGCAGCTTTTCTGCTGCAGA 100
|||||
321 pAspArgLeuAsnArgMetAsnGlnAlaSerArgLeuPheLeuLeuGlnT 338
101 CAGTTGAGCTGGGTCCTGGGTTGGGATGGTGCATTTGACAGTGCCTGCC 150
|||||
338 hrValGluGlnGlySerTrpValGlyMetValThrPheAspSerAlaAla 354
151 CATGTACAAAGTGAATCATACAGATAAAGCAGGCGCAGTGCAGGGACAC 200
|||||
355 TyrValGlnSerGluLeuLysGlnLeuAsnSerGlyAlaAspArgAspLe 371
201 ACTCGCCAAAAGATTACCTGCAGCAGCTTTCAGGAGGAGCGCTCCATCTGCA 250
|||||
371 uLeuIleLysHisLeuProThrValSerAlaGlyGlyThrSerIleCysS 388
251 GC 252
388 er 388
```

seq_name: /SID52/gcgdata/geneseq/AA2001.DAT: AAB73715

seq_documentation_block:

ID AAB73715 standard; protein: 913 AA.

XX AC AAB73715;

DT 11-SEP-2001 (first entry)

XX DE Mouse Gob-5 protein, SEQ ID NO:1.

XX KW Mouse Gob-5; murine; goblet cell; human CLCA1 orthologue; drug screening; expression inhibition; antisense therapy; gene therapy; bronchial asthma; chronic obstructive pulmonary disease; antiasthmatic.

XX OS Mus sp.

XX PN WO200138530-A1.

XX PD 31-MAY-2001.

XX PF 22-NOV-2000; 2000WO-JP08232.

XX PR 24-NOV-1999; 99JP-0333479.

XX PR 27-APR-2000; 2000JP-0127589.

XX PA (TAKE) TAKEDA CHEM IND LTD.

XX PI Nakanishi A, Morita S;

XX WPI; 2001-355935/37.

DR N-PSDB; AAH46101, AAH46120.

XX PT New antisense nucleotide, useful for treatment and prevention of

XX PT bronchial asthma and chronic obstructive pulmonary disease

XX PS Claim 1; Page 72-76; 104pp; Japanese.

XX CC The invention relates to an antisense nucleotide targetted to the mouse Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart, the CLCA1 gene (coding sequence shown in AAH46102). The invention also relates to an antibody specific for the Gob-5 protein, medical and diagnostic compositions containing the antisense nucleotide or the antibody, and methods and kits for screening for compounds which inhibit the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells. The antisense oligonucleotides and antibody are therefore useful for the treatment and prevention of bronchial asthma and chronic obstructive

CC pulmonary disease. The present sequence represents mouse Gob-5 protein.

XX SQ Sequence 913 AA;

XX alignment_scores:

Quality: 330.00 Length: 84

Ratio: 4.231 Gaps: 0

Percent similarity: 92.857 Percent identity: 77.381

alignment_block:

US-09-049-696-6 x AAB73715

Align seg 1/1 to: AAB73715 from: 1 to: 913

1 CAAAGAAATGTTGTTAGTCTTGACAAATCTGGAAGCATGGCGACTGG 50

|||||

305 GlnArgIleValCysLeuValLeuAspLysSerGlySerMetLeuAsnAs 321

51 TAACCGCCTCAATCGACTGAATCAAGCAGCGCAGCTTTTCTGCTGCAGA 100

|||||

321 pAspArgLeuAsnArgMetAsnGlnAlaSerArgLeuPheLeuLeuGlnT 338

101 CAGTTGAGCTGGGTCCTGGGTTGGGATGGTGCATTTGACAGTGCCTGCC 150

|||||

338 hrValGluGlnGlySerTrpValGlyMetValThrPheAspSerAlaAla 354

151 CATGTACAAAGTGAATCATACAGATAAAGCAGTGCAGTGCAGGGACAC 200

|||||

355 TyrValGlnSerGluLeuLysGlnLeuAsnSerGlyAlaAspArgAspLe 371

201 ACTCGCCAAAAGATTACCTGCAGCAGCTTTCAGGAGGAGCGCTCCATCTGCA 250

|||||

371 uLeuIleLysHisLeuProThrValSerAlaGlyGlyThrSerIleCysS 388

251 GC 252

388 er 388

seq_name: /SID52/gcgdata/geneseq/AA2000.DAT: AAY66749

seq_documentation_block:

ID AAY66749 standard; protein: 919 AA.

XX AC AAY66749;

XX DT 05-APR-2000 (first entry)

XX DE Membrane-bound protein PRO1124.

XX KW Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand; pharmaceutical; receptor immunoadhesin; gene mapping.

XX OS Homo sapiens.

XX PN WC9963088-A2.

XX PD 09-DEC-1999.

XX PF 02-JUN-1999; 99WO-US12252.

XX PR 02-JUN-1998; 98US-0087607.

XX PR 02-JUN-1998; 98US-0087609.

XX PR 02-JUN-1998; 98US-0087759.

XX PR 03-JUN-1998; 98US-0087827.

XX PR 04-JUN-1998; 98US-0088021.

XX PR 04-JUN-1998; 98US-0088025.

XX PR 04-JUN-1998; 98US-0088028.

XX PR 04-JUN-1998; 98US-0088029.

XX PR 04-JUN-1998; 98US-0088030.

XX PR 04-JUN-1998; 98US-0088033.

XX PR 04-JUN-1998; 98US-0088326.

XX PR 05-JUN-1998; 98US-0088167.

PR 05-JUN-1998; 98US-0088202.
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 26-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.

PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 18-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 31-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.

(GETH) GENENTECH INC.

Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
Wood WI, Yuan J;

WPI; 2000-072883/06.
N-PSDB; AA265095.

Membrane-bound proteins and related nucleotide sequences -
claim 12; Fig 274; 822pp; English.

The invention provides membrane-bound PRO polypeptides and polynucleotides encoding them. The PRO sequences of the invention were identified based on extracellular domain homology screening. The PRO sequences have homology with proteins including LDL receptors, TIE ligands and various enzymes. The membrane-bound proteins and receptor molecules are useful as pharmaceutical and diagnostic agents. Receptor immunoadhesins, for instance, can be used as therapeutic agents to block receptor-ligand interactions. The membrane-bound proteins can also be employed for screening of potential peptide or small molecule inhibitors

CC of the relevant receptor/ligand interaction. The PRO encoding sequences
CC are useful as hybridization probes, in chromosome and gene mapping and in
CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC will also be useful for the preparation of PRO polypeptides, especially
CC by recombinant techniques.
XX
SQ Sequence 919 AA;

alignment_scores:
Quality: 268.00 Length: 84
Ratio: 3.775 Gaps: 0
Percent Similarity: 84.524 Percent Identity: 66.667

alignment_block:
US-09-049-696-6 x AAY66749

Align seg 1/1 to: AAY66749 from: 1 to: 919

```
1 CAAGAATTGCTGTTAGTCTTGACAAATCTGGAAGCATGGCGACTGG 50
|||||
304 GlnArgIleValCysLeuValLeuAspLysSerGlySerMetGlyGly 320
51 TAACCCCTCAATCGACTCAATCAAGCAGCCAGCTTTCTCCCTCCGAGA 100
|||||
320 sAspArgLeuAsnArgMetAsnGlnAlaAlaLysHisPheLeuLeuGln 337
101 CAGTTCAGCTGGGTCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 150
|||||
337 hrValGluAsnGlySerTrpValGlyMetValHisPheAspSerThrAla 353
151 CATGTACAAGTGAATCATACAGATAACAGTGGCAGTGCAGGAGGACAC 200
|||||
354 ThrIleValAsnLysLeuIleGlnIleLysSerSerAspGluArgAsn 370
201 ACTCGCCAAAGATTACCTGACGAGCTTCAGGAGGAGCTGCATCTGCA 250
|||||
370 rLeuMetAlaGlyLeuProThrTyrProLeuGlyGlyThrSerIleCys 387
251 GC 252
||
387 er 387
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB87560

seq_documentation_block:
ID AAB87560 standard; Protein; 919 AA.

```
XX
AC AAB87560;
DT 15-MAY-2001 (first entry)
DE Human PRO1124.
KW Human; PRO protein; mapping.
XX Homo sapiens.
OS
XX WO200116318-A2.
PN
PD
XX
XX 08-MAR-2001.
XX 24-AUG-2000; 2000WO-US23328.
XX 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 07-DEC-1999; 99US-0169495.
PR 09-DEC-1999; 99US-0170262.
PR 11-JAN-2000; 2000US-0175481.
PR 18-FEB-2000; 2000WO-US04341.
PR 22-FEB-2000; 2000WO-US04342.
PR 01-MAR-2000; 2000WO-US05601.
```

PR 03-MAR-2000; 2000US-0187202.
PR 25-APR-2000; 2000US-0199397.
PR 22-MAY-2000; 2000WO-US14042.
PR 05-JUN-2000; 2000US-0209832.
XX
PA (GETH) GENENTECH INC.

XX Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
XX
XX WPI: 2001-183260/18.
DR N-PSDB; AAF92092.

XX Eighty four nucleic acids encoding PRO polypeptides, useful in
PT molecular biology, including use as hybridization probes, and in
PT chromosome and gene mapping.

XX
PS Claim 12; Fig 70; 278pp; English.

XX The present sequence is a human PRO polypeptide (secreted and
CC transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
CC anti-PRO antibodies are useful for preparation of a medicament useful in
CC the treatment of a condition which is responsive to the PRO protein,
CC agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
CC employed as molecular weight markers for protein electrophoresis. The PRO
CC coding sequence has applications in molecular biology, including use as
CC hybridisation probes, and in chromosome and gene mapping.

XX
SQ Sequence 919 AA;

alignment_scores:
Quality: 268.00 Length: 84
Ratio: 3.775 Gaps: 0
Percent Similarity: 84.524 Percent Identity: 66.667

alignment_block:
US-09-049-696-6 x AAB87560

Align seg 1/1 to: AAB87560 from: 1 to: 919

```
1 CAAGAATTGCTGTTAGTCTTGACAAATCTGGAAGCATGGCGACTGG 50
|||||
304 GlnArgIleValCysLeuValLeuAspLysSerGlySerMetGlyGly 320
51 TAACCCCTCAATCGACTCAATCAAGCAGCCAGCTTTCTCCCTCCGAGA 100
|||||
320 sAspArgLeuAsnArgMetAsnGlnAlaAlaLysHisPheLeuLeuGln 337
101 CAGTTCAGCTGGGTCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 150
|||||
337 hrValGluAsnGlySerTrpValGlyMetValHisPheAspSerThrAla 353
151 CATGTACAAGTGAATCATACAGATAACAGTGGCAGTGCAGGAGGACAC 200
|||||
354 ThrIleValAsnLysLeuIleGlnIleLysSerSerAspGluArgAsn 370
201 ACTCGCCAAAGATTACCTGACGAGCTTCAGGAGGAGCTGCATCTGCA 250
|||||
370 rLeuMetAlaGlyLeuProThrTyrProLeuGlyGlyThrSerIleCys 387
251 GC 252
||
387 er 387
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB65272

seq_documentation_block:

ID AAB65272 standard; Protein; 919 AA.

```
XX
AC AAB65272;
XX 02-APR-2001 (first entry)
DT
```


DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF00092; vwa; 1
DR Pfam: PF00097; zf-C3HC4; 1.
DR SMART: SM00184; RING; 1.
DR SMART: SM00327; VWA; 1.
DR PROSITE: PS50234; VWFA; 1.
KW Zinc-finger.
SQ SEQUENCE 689 AA; 75704 MW; 96DF79BE29CE5CF2 CRC64;

alignment_scores:
Quality: 153.00 Length: 256
Ratio: 0.987 Gaps: 12
Percent Similarity: 60.547 Percent Identity: 26.562

alignment_block:
US-09-049-696-20 x Q9ZV10 ..
Align seg 1/1 to: Q9ZV10 from: 1 to: 689

952 GTCCTTGACAAATCTGGAGCATGCGACTGTAACCGCTCAATCGACT 1001
AC Q9PLZ9; PRELIMINARY; PRT; 928 AA.
254 ValLeuAspIleSerGlySerMetAla...GlyThrLysLeuAlaLeuLe 269
1002 GAATCAAGCAGCCAGCTTTCTCTGCTGCGACAGATTGAGTGGGGTCT 1051
GN CJ1678.
269 ulysArgAla...MetGlyPheValIleGlnAsnLeuGlySerAsnAspA 285
1052 GGGTTGGGATGTCGATTTGACATTTGACATGTCGCCATGTACAAAGTCAATC 1101
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
285 rgLeuSerValIleAlaPheSerThrAlaArgLeuPheProLeu 301
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
1102 ATACAGATAAAGTGGCAGTGGACAGACACACTGCCAAAGAAAGATTACC 1151
RA Quail M.A., Rajadream M.A., Ruthford K.M., Van Vliet A.H.M.,
302 ThrLysMetSerAspAlaGlyArgGlnArgAlaLeuGlnAlaValAsnSe 318
1152 TGACAGAGCTTCAGGAGGACGTCATCTGACGGGCTTCATCGGCAT 1201
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
318 rValValAlaAsnGlyThrAsnIleAlaGluGlyLeuArgLysGlyV 335
DR EM3L; AL139079; CAB73665.1; -.
1202 TTACTGTGATT.....AGGAAGAAATATCCAACTGATGATCTGAA 1242
KW Complete proteome.
335 allysValMetGluAspArgArgAspLysAsnProVal.....AlaSer 349
1243 ATTGTGCTGCTGACGATGGGAGACAC..... 1272
350 IleIleLeuLeuSerAspGlyArgAspThrTyThrMetAsnGlnAlaAs 366
1273ACTATAAGTGGTGTCTTTAACG 1294
366 pProAsnTyLysLeuLeuLeuProLeuSerMetHisGlyCys.....G 381
1295 AGGTCAACAAAGTGTGTCATCATCCACACATCGCTTTGGGGCCTCT 1344
381 luSerLysArgPheGlnIleProValHisSerPheGlyPheGlySerAsp 397
1345 ...GCAGCTCAAGACTAGAGGAGTGTCCAAATGACAGGAGGTTTACA 1391
398 HisAspAlaSerLeuMetHisSerValSerGluThrSerGlyGlyThrPh 414
1392 GACATAT...GCTTCAGATCAAGTTCAGAACATGGCCCTCATTTGATGCTT 1438
414 eSerPheIleGluSerGluSerValIleGlnAspAlaLeuAlaGlnCysI 431
1439 TTGGGGCCCTTTTCATCAGGAATGGAGCTGTCTCTCAGCCCTCCATCCAG 1488
431 leGlyGlyLeuLeuSer.....ValAlaValGlnGluLeuArg 443
1489 CTTGAGAGTAAGGATTA..... 1506
444 LeuGluIleGluGlyMetCysSerAspValHisLeuSerSerIleLysAl 460
1507ACCCTCCAGAACAGCCAGTGGATGAATGGCACAGTG. 1542

460 aglySerTyrglnSerLeuValserGlyArgSerGlyCysVala 477
1543 ..ATCGTGGACACGCGTGGGAAAGACACTTTGTTCTTATCACCTGG 1590
477 spIleGlyAspLeuTyAlaAspGluGluArgAspPheLeuSer.... 492
1591 ACAACGACGCTCCCA 1608
493 ValAsnIleProGln 498

seq_name: sp_bacteria:Q9PLZ9
seq_documentation_block:
ID Q9PLZ9 PRELIMINARY; PRT; 928 AA.
AC Q9PLZ9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE POSSIBLE LIPOPROTEIN (FRAGMENT).
GN CJ1678.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCCT 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajadream M.A., Ruthford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EM3L; AL139079; CAB73665.1; -.
KW Complete proteome.
FT NON_TER 1
SQ SEQUENCE 928 AA; 97589 MW; 5A724384245C3CAB CRC64;

alignment_scores:
Quality: 151.50 Length: 913
Ratio: 0.384 Gaps: 42
Percent Similarity: 43.264 Percent Identity: 18.401

alignment_block:
US-09-049-696-20 x Q9PLZ9 ..

Align seg 1/1 to: Q9PLZ9 from: 1 to: 928

298 GTGAGACCAAACTTGACACCTACAAAATGCTGATGTTCTGGTTGCTGA 347
50 IleAsnThrLysIleAspThrPheThrAsnAsnGlyPheIleAsnSerPr 66
348 GTCCTACTCCTCCAGGTAATGATGAACCTCAGCTGACGATGGGCAACT 397
66 oclySerGlyGlnTrpAsnAsnGlyIleTrp.IleSerSerAsnAlaThr 82
398 GTGGAGAGAGGGTGAA.....GGATCCACTCACT 429
83 IleGluLysLeuValAsnAsnGlyThrIleLysGlyGlyHisSerAlaI 99
430 CTTGATTTTCATTCAGGAGAAAGTTAGCTGAATATGCACCACAGGTAG 479
99 eMetValThrSerGlnHis.....IleL 107
480 GGCATTGTCATGATGAGTGGCTCATCTACGA.....TGGGA... 516
107 ysThrValGluAsnThrGlyIleIleHisAlaGluGlyGluTrpGlySer 123

517 ...GTATTCACGACGACATAATATGATGACAAATTCCTACTTATCCCAATGG 563
124 SerIleuLeuGluThrGlyPheIleGluHisIleIleAsnThrGI 140
564 AAGAATACAA.....GCAGTAAGATGTTCCAGCAGGTATTACTG 601
140 yThrIleSerAsnAsnValGlyIleGlySerAlaTyrGlyValPheG 157
602 GTACAAATGAGTAAGAAGTGTCCAGGAGGACGCTGTACACCAAA... 648
157 lyThrIleIleLys.....AspGlyGlyMetValTyrGlyLysTyr 171
648 648
172 SerAlaIleGlyValGlyArgSerGlnThrLeuGlyAspLeuTyrIleAs 188
649AGATGCACATTCATTAATAAGTAACAGGACTCTATGAAAAAGGATGTG 694
188 pGlyArgSerAsnAsnGlyThrValSerGlyIleTyrSerGluGluHisG 205
695 AGTTTGTCTCCAAATCCCGCCAGCAGGAGAAAGGCTCTATAATGTTTGA 744
205 lyIleLeuLeuGluAsn..... 210
745 CAACATGTTGATTCTATAGTTGAATTCGTACAGAAACAAACACACAA 794
210 210
795 AGAAGCTCCAAACAAAGCAAAATCAAAATGCAATCCGGAAGCACATGGG 844
211AsnSerArgThrGlnIleGluLeuLysAsn...GlyC 223
845 AAGTGATCCGTGATCTGAGGACTTTAAGAAACCCTCTATGACAAACA 894
223 lyIleIleLysGlyAsnIleAsp..... 230
895 CAGCCACCAATCCACCTTCTATTGCTGTCAGATTGGACAAAGAATTGT 944
231GlyIleArgLeuII 235
945 GTGT.....TAGTCCTTGACAAATCTGGAAGCA 973
235 eAsnSerAlaSerLeuSerGlyGluMetIleLeuSerGlyGluGlySerA 252
974 TGGCGACTGTAAACCG...CTCAATGACTGAATCAAGCAGGCCAGCT 1020
252 rgValGluGlyGlyArgGlyValGlyIleLeuAsnArgSerGlyLysIle 268
1021 TTCTCTCTGCAGACGTTGAGCTGGGTCTCTGCTGTT..... 1056
269 GluGlySerIleLysValGluAspGlyAlaThrValThrAlaThrSerAs 285
1057GGGATGGTGACA..... 1068
285 nArgAlaIleAlaAsnSerGlySerGlySerIleThrGlyGlyIleThrV 302
1068 1068
302 alSerGlyLysAsnThrLysLeuGluGlyAsnIleIleAsnThrGlyAsn 318
1069TTTGACAGTGTGCCCATCTACAAAG 1094
319 AlaSerIleGlySerAspIleLysIleGluGlyGlyAlaLysValGluGI 335
1095 TGAATCT.....ATACAGATAA 1111
335 yGlyLeuValAsnGlnGlyAsnGlySerIleSerGlySerValGlnValS 352
1112 ACAGTGGCTGACAGGACACACTCGCCAAAAGATTACTCGCAGCAGCT 1161
352 erglyGlySerSerIleAspSerIleThrAsnGluGlyAsnGlyAlaIle 368
1162 TCAGGAGGACGTCCTATCTGC.....ACGGGCTTCGATCGGCATTAC 1205

369 SerGlySerIleThrValTyrLysAspSerLysLeuAspSerIleThrAs 385
1206 TGTGATTAGGAAGAAATATCCAACTGATGATCTGAAATTTGTCTGCTG 1255
385 nThrSerThrSerThrGlyIleSerGlySer.....IleT 398
1256 CGGATGGGAAGACAAACACTATTAAGTGGGTCTTTAACGAGGTCAACAA 1305
398 hrAsnAsnSerAspAsnLysLeu.....GluIleSerAsn 409
1306 AGTGGTCCCATCATCCACAGTCGCTTTGGGGCCCTCTCCACGCTCAAGA 1355
410 SerGlyAsnIle..... 413
1356 ACTAGAGGAGCTGTCCAAAATGACAGGAGGTTTA.....CAGACATATG 1399
413GlyGlyLysIleGluSerThrGlyS 422
1400 CTTTCAGATCAAGTT.....CAGAACAAATGGCTCATTTGATGCTTTT 1440
422 erAlaAspMetValIleSerAsnSerAsnGlyGlyThrIleSerGly... 437
1441 GGGGCCCTTTTCATCAGAAATGGAGCTGTCTCTCAGCGCTCCATCCAGCT 1490
438 GlyIleSerSerGlySerGlySerThrSerIleSerAsn..... 451
1491 TGAGAGTAAGGATTAAACCTCCAGAACAGCCAGTGGATGAATGGCACAG 1540
452SerGlnGlySerThrIleAsnAsnGlyIleThrValSerGlySerA 467
1541 TGATCTGTGGAC.....AGCACCTGGGAAAGGACACTTTGTTTT 1578
467 laglnValGluIleSerAsnGlnGlySerValGlyLysAsp..... 480
1579 CTTATCACCTGGACAACGAGCGCTCCCAAAATCCTCTCTGGGATCCAG 1628
480 480
1629 TGGACAGAAGCAAGTGGCTTTCTAGTGGACAAAACACCAAAATGGCCT 1678
481GluAsnGlyAsnThrValThrAsnAsn..... 489
1679 ACCTCCAAATCCAGGACATTCCTAAGTTGGC.....ACTTGGAAATAC 1722
490GlySerGlySerValGlyIleLysAspTrpLeuVal 501
1723 AGTCTGCAAGCAAGCTCACAAACCTTGACCTGACTGCTCAGC.....TC 1766
502 SerThrAspLysAsnThrGlyLysLeuAsnThrValValIleGlyGlySe 518
1767 CCGTGGTCCAATGCTACCTGCTCCAAATACAGTGACTTCCAAAACGA 1816
518 rArgAlaPheAsnValLysValGluAsnIleThrVal...AspGlnSerA 534
1817 ACAAGCACACCACCAAAATCCCCAGCGCTCTGGTAGTTTATGCAAAATAT 1866
534 snValAspLeuGluLeu..... 540
1867 CGCAAGGAGGCTCCCAATTCCTCAGGGCCAGTGTCACAGCCCTGATTGA 1916
540AsnAspIleAsnAsnIleIleSe 548
1917 ATCAGTGAATGGAAAACAGTTTACCTTGGAACTACTGGATTAATGGAGCAG 1966
548 rGlyValAsnGlnAsnAsnIleGly...AsnIleGlyThrAsnGlySerG 564
1967 GTCTGATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCACAACT 2016
564 lyIleSerLeuSerPheAspProIleThrGlyLysLeuThrThrAsp 580
2017 TATGACACGAATGATACATACGT...GTAAAAGTCGGGCTCTGGGAGG 2063

```

581 PheAsnLeuAsnAlaSerIleSerGlyAlaThrPheArgSerLeuIleSe 597
2064 AGTTAACGACGACGACGAGGAGTGATACCCAGCAGAGTGGAGCACTGT 2113
      :::::  |||
597 rThrThrSerArgSerThrPheIle..... 606
2114 ACATACCTGGCTGGATGAGAAATGAGAAATACAAATGGAATCCACCAGA 2163
      ||:::  |||
607 .....AspAsnValMetGlyAsn..... 612
2164 CCTGAATTAATAGGATGATCTTCAACACAGCAAGTGCTGTTTCAGCAG 2213
      :::::  :::::
613 .....SerMetGlnSerPheAlaLeuAlaSe 621
2214 AACATCTCGGGAGCTCATTTGTGCTTCATGTCGCCAAAT.....G 2257
      :::::  ||:::  ||:::
621 rSerSerLysSerGlnSerIleAlaMetSerGlnLysGlyAsnLeuTyra 638
2258 CTCCCATACCTGATCTCTCCACCTGCCAAATCACCAGCCTGAGGCG 2307
      ||  |||
638 laAspAlaSerAspTyr.....IleLysSer 646
2308 GAAATTCAGGGGCGAGTCTCATTAATCTGACTTGGACAGCTCCTGGGGA 2357
      ||:::  ||:::
647 AspLeuAsnAsnGlySer..... 652
2358 TGATTATGACCATGGAACAGCTCACAAGTATATTCGATAAGTACAA 2407
      |||  :::::
653 ....TyrGlySerAsnLysGluHisSerLeuPheIleLeuProTyrThrS 668
2408 GTATTCTTCATCTCAGACAGATGTCATGATCTCTCAAGTGAATACT 2457
      ||  :::::  ||:::  ||:::
668 exSerGlnAsnValGluLeuSerLeuAsnGluLysSerLysGlyHisThr 684
2458 ACTGCTCTCATC.....CCAAAGGAAGCCAACTCTCAGGA 2492
      :::::  |||
685 LysGlyThrIleIleGlyTyrSerThrLeuLysAspSerGlyIleTyrGl 701
2493 AGTCITTT.....TTGTTTAAAC 2509
701 yValTyAlaGlyTyTyGluAspThrLysMetGlySerThrTyPheAspI 718
2510 CAGAAACATTACTTTT..... 2526
718 leAsnAsnArgThrTyTyAlaGlyLeuLysTyTyPheAsnThrLeuPhe 734
2527 .....GAAATGGCACAGATCTTTTCATGCTATTCAG.....GC 2561
      ||:::  |||  :::::  |||
735 ThrThrGluLysGlyGlnGluValTyIleLysAlaGlnGlyLysAlaAl 751
2562 TGTGTGATAAGTCGATCTCAATCAGAAATATCCAAAC 2598
      ||:::  |||  ||:::  ||:::  ||:::
751 aleuIleLysAsnAspLeuThrGluLysIleGlyAsn 763

```

seq_name: sp_bacteria:Q9L950

```

seq_documentation_block:
ID Q9L950 PRELIMINARY; PRT: 2147 AA.
AC Q9L950;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE OUTER MEMBRANE-LIKE PROTEIN.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20225830; PubMed=10762233;
RA Espinosa-Urgel M., Salido A., Ramos J.L.;
RT "Genetic analysis of functions involved in adhesion of Pseudomonas
putida to seeds.";

```

```

RL J. Bacteriol. 182:2363-2369(2000).
RN [2].
RP SEQUENCE FROM N.A.
RA Espinosa-Urgel M., Salido A., Ramos J.L.;
RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF182515; AAF69025.1;
DR InterPro: IPR001343; Hemlysn_Ca_bind.
DR PRINTS: PR00313; CABDNRPRT.
SQ SEQUENCE 2147 AA; 211585 MW; BD7FC3B460FE2D6B CRC64;

```

```

alignment_scores:
      Quality: 149.50      Length: 774
      Ratio: 0.430      Gaps: 37
Percent Similarity: 44.961 Percent Identity: 19.897
alignment_block:
US-09-049-696-20 x Q9L950 ..
Align seg 1/1 to: Q9L950 from: 1 to: 2147

```

```

844: GAAGTGATCCGTGATCTGAGGACTTTAAAGAACCACTCTCTATGACAAC 893
      :::::  |||  :::::
337 GlnValSerAlaThrGlyProAspGlyThrValSerLeuProAlaThrLe 353
894 ACAGCCACCAAAATCCACCTTCTCATTGCTGCAGATTGGACAAAGAAATTG 943
      |||  ||:::  |||  :::::
353 uGlnAlaProAspThrThr.....AlaProGlnProLeuT 365
944 TGTGTTTAGTCTTGACAAATCTGGAAGCATGCGAGCTGTTAAC..... 987
      :::::  ||:::  ||:::  ||:::
365 hrAsnLeuValLeuSerAsnAspGlyLeuValLeuSerGlyArgGlyGlu 381
988 .....CCCTCAATCGACTCAATCAACGACGCGCCAGCTTTTCCTGCT 1028
      :::::  ||:::  ||:::  ||:::
382 AlaGlyAlaThrValThrValLeuGlyGluGlyGly..... 393
1029 GCAGACAGTTGAGCTGGGCTCTGGTGGATGGTGACATTTGACAGTG 1078
      :::::  ||:::  ||:::  ||:::
394 .....IleGluLeuGlyThr.....AlaValValAspAlaAsnGlyA 406
1079 CTGCCCCATGACAAAGTGAAGTCACTATACAGATAAAC..... 1113
      :::::  ||:::  ||:::  ||:::
406 spPheAsnValAspLeuSerValAlaGlnIleAsnGlyGluGlnLeuSer 422
1114 .....AGTGGCAGTGACAGGACACACTCGCCAAAAG 1145
423 AlaSerGlnSerAspValAlaGlyAsnGluSerAsnSerValSerLeuTh 439
1146 ATTACCT.....GCAGCAGCTTCAG 1165
      |||  |||  |||
430 rAlaProAspArgValAlaProAspAlaAlaGlyAsnLeuAlaPheAlaG 456
1165 GAGGAGCTCCATCTGCGAGCGGCTTCGATCGGCATTTACTGTGATAGG 1215
      ||:::  ||:::  ||:::  ||:::
456 lyGlyGlySerLeuLeuAsnGlyThrGlyGluAlaGlyAlaAlaValArg 472
1216 AAGAAATATCAACTGATGATCTGAAATTTGCTGCTGCGGATGGGA 1265
      ||:::  ||:::  ||:::
473 .....ValLeuAlaAlaAsp..... 477
1265 AGACAACACTATAAGTGGTCTTTAACAGAGGTCAACAAAGTGGTGCCA 1315
      ||:::  ||:::
478 .....GlySerV 480
1315 TCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGGAG 1365
      :::::  :::::
480 alLeuGlySerAlaThrValArgAlaAspGlyThrPheGlnValAsnLeu 496
1366 CTGTCCAAATGACAGCAGGT.....TTACAGACATATGCTTCAGATCA 1409
      :::::  ||:::  ||:::
497 ValSerProGlnAlaAsnGlyGlnGlnValValLeuAlaAspAl 513

```

```
1410 AGTTCAGACAATGGCTCTATTGATGCTTTGGGGCC..... 1446
... :|||:||||:|
513 aAlaGlyAsnGlnSerAlaAlaSerAlaIleThrAlaProAspThrThrp 530
1447 .....CTTTCATCAGAAATGGAGCTCTCTCAGCGCTCCATCAGCTT 1491
... :|||:|||||
530 roProSerAlaProGlyGluLeuAlaValSerGlyAspGlyValThrLeu 546
1492 GAGAGTAAGGA.....TTAACCTCCAGAACACCCAGTG 1526
... :|||:|
547 ThrGlyArgGlyGluValGlyAlaArgValThrValThrAsnProAla.. 562
1527 GATCAATGGCAGCTGATCGTGACACACCGTGGGAAAGCAGACACTTGT 1576
563 .....GlyGluGluIleGlySerAlaThrValAspAlaSerGlyHisP 577
1577 TTCTTATCACTGG..... 1590
577 heThrIleThrLeuSerProAlaLeuGlyAsnAlaGluLeuLeuSerPhe 593
1591 .....ACAGCGAGCC 1601
594 ThrGlnAlaAspAlaAlaGlyAsnValSerAlaValAlaIleThrLeuGlnTh 610
1602 TCCCAAAATCCTCTCTGGGATCCAGTGGACAG..... 1635
610 rProAspPheThrProProAspProLeuThrGlnValValIleAsnAla 627
1636 .....AAGCAAGGTGCTTTAGTGGAC 1659
627 spGlySerValValSerGlyLeuGlyGluAlaGlyAlaThrValIleVal 643
1660 AAAACACCAAAATGGCTACCTCCAATCCAGGCATGCTAGGTGG 1709
644 SerAsnAlaAlaGlyThrValLeuGlyThrGlyIleValLeuAlaAspG1 660
1710 CACTTGGAAATACACTGTG.....CAAGCAAGCTCACAACCTTGA 1750
660 ySerPheSerValGluLeuSerProAlaGlnIleAsnAsnGlnThrLeuS 677
1751 CCTGTACTGTACGTCCTCGGTCGCTCAATGCTACCCCTG..... 1788
677 ervAlLeuGlnAlaAspProProGlyAsnValSerLeuProValAsnVal 693
1789 .....CCTCCAATTACAGTGACTCC...AAACGAA 1817
694 GlnAlaProAspLeuThrProAlaAlaAlaThrGluLeuArgLeuAs 710
1818 CAAG.....GACACCAGCAATTCGCCAGCCCTCTGG 1849
710 nAlaAlaGlyAspGluLeuGlyGlyValGlyGluValGlyAlaThrValA 727
1850 TAGTTTATGCAATATTTCGCCAAGGAGGCTCCCAATTCACAGGCCAGT 1899
727 rgValTyValGlyThrGlnIleGlyThrGlyThrValGlyAlaAsn 743
1900 GTCACAGCCCTGATTGAA.....TCAGTGAATGGAACACAGT 1937
744 GlyThrPheIleValGluLeuAspAlaAlaGlnLeuAspGlyGlnValLe 760
1938 TACCTTGGAACTACTGATTAATGGCAGCAGGTGCTGATGCTACTAAGGATG 1987
760 uGlnValThrLeu.....ThrAspAlaThrGlyAsnI 771
1988 ACGGTGTCTACTCAAGTATTTCCAACTTATGACACGAAATGGTAGATAC 2037
771 leSerProIleSerSer...ValThrAlaValAlaAspThrThrProAla 786
2038 AGTGTA.....AAAGTCGGGCTCTGGGAGGAGT 2066
787 ThrValValAlaThrIleAsnAspSerGlyThrGlnValIleGlyThrAl 803
2067 TAACCGACCCAGCGGAGAGTGTATACCCAGCAGCAGAGTGGAGCACTGTACA 2116
```

```
803 aGluAlaGlyThrArgValThrValValAsnAsnGlyAspLeuLeuG 820
2117 TACCTGGCTGGATTGAGATGAT.....GAAATACAAATGAATCCA 2157
820 lyGlnValThrValGluProAspGlyThrPheValLeuAspLeuAsn... 835
2158 CCAAGACCTGAAATTAATAAGGATGATGTTCAACACAAAGCAAGTGTCTTT 2207
836 ...LeuProGlnIleAsnGlyGluThrLeu..... 844
2208 CACGAGAACATCCTCGGAGGCTCATTGTGGCTTCTGATGTC..... 2250
845 .....ThrIleIleAlaGlnAspAlaThrGlyA 854
2251 ...CCAAATGCTCCATA.....CCTGATCTCTCCACCTGGC 2286
854 snProSerValProLeuAlaValThrAlaProAspLeuThrProVal 870
2287 CAATCACCGACCTGAAGCGGAAATTCACGGGGCAGCTCATTAATCT 2336
871 GlnProThrAspLeuLeuAsnAlaAlaGlyThrSerMetThrGlyTh 887
2337 G.....ACTTGGACAGCTCCTGGGATGAT..... 2361
887 rAlaGluAlaGlySerThrIleThrValArgGlyProAspAsnIleV 904
2362 .....TATGACCATGGACAGCTCACAAGTATATCATTCGAATAAGT 2403
904 aGlyThrValGluValGlyGluAsnGlyGlnPheThrValProIleSer 920
2404 .....ACAAGTATCTTGTGATCTCAGAGACAA 2429
921 ProAlaGlnAsnAsnGlyGlnLeuLeuThrValValSerThrAspAspAl 937
2430 GTTCAATGAATCTCTT.....CAAGTGAATACTACTGCT... 2463
937 aGlyAsnAlaSerValAlaValProTyThrGlnThrAsnAspThrSerProp 954
2464 ....CTCATCCCAAGGAGGCC.....AACTCTGAGGAGTCTTTTGTG 2502
954 roAspLeuValThrAsnLeuAlaIleAsnAsnThrPheAsnValLeuThr 970
2503 TTAAACACAGAA.....AACATTACTTTTCAAAAT..... 2532
971 GlyGlyGlyGluAlaAsnAlaThrValThrValSerPheAspGlyAspVa 987
2533 ...GCCACA.....GATCTTT 2545
987 lIleGlyThrGlyValValAspGlyAlaGlyAsnPheGlnIleThrLeuV 1004
2546 TCATTGCTATTTCAGGCTGTTGTGATAAGGTCGATCTG...AAATCAGAAATA 2592
1004 alProAlaValGlySerLeuAlaThrLeuGluValThrGlnSerAspGly 1020
2593 TCCAACATTCACAGCATCTTTGTTTATCTCCACAGACTCCGCCAGA 2642
1021 IleAsnThrSerAlaValAlaThrTyThrLeuThrProLeuIleAlaProPr 1037
2643 GACACCTAGTCTGATCAACAGC 2664
1037 ovalProThrAspValThr 1044
seq_name: sp_bacteria:Q9AES2
seq_documentation_block:
ID Q9AES2 PRELIMINARY; PRT: 1234 AA.
AC Q9AES2;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE IGA PROTEASE.
GN IGA.
```

OS Clostridium ramosum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes.
 OX NCBI_TaxID=1547;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AK183;
 RA Kosowska K.; Reinholdt J., Kjaer L., Potempa J., Kilian M.,
 RA Poulsen K.;
 RT "The Clostridium ramosum immunoglobulin A (IgA) protease represents a
 RT novel metalloproteinase."
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY028440; AAK27311.1; -
 KW Protease.
 SO SEQUENCE. 1234 AA; 137182 MW; A4BFE42A2A34A87 CRC64;

alignment_scores:

Quality: 144.50 Length: 757
 Ratio: 0.450 Gaps: 35
 Percent Similarity: 42.404 Percent Identity: 19.419

alignment_block:

US-09-049-696-20 x Q9AES2 ..

Align seg 1/1 to: Q9AES2 from: 1 to: 1234

```

355 CTTCCAGGTAATGAT...GAACCTACTAGCAGAGTGGGCACTGTGG 401
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
482 ProSerGlySerGluTyrGluProTyrTyrValHisAspTyrIleAl 498
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
402 AGAGAGGCTGAAGGATCCACCTCCTCATGATTCATTCAGGAAAAA 451
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
498 aGlnPheAlaMetValValAsnThrLysSerAspPheGlyGlyAlaTyrA 515
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
452 AGTTAGCTGAATATGGA..... 468
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
515 snAsnArgGluTyrGlyPheHisTyrPheIleSerProSerAspSerTyr 531
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
469 CCACAGGTAGGCGATTTCCATGAGTGGGCTCATCAGTGGGAGT 518
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
532 ArgAlaSerLysThrPheAlaHisGluPheGlyHisGlyLeuLeuGlyLe 548
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
519 ATTTCAGGAGTACAATAATGATGAGAAATCTACTATCAATGGAAGAA 568
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
548 uGlyAspGluTyrSerAsnGly.....TyrLeuLeuAspAspLysGluL 563
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
569 TACAAGCAGTACAGTGTTCAGCA..... 591
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
563 euLysSerLeuAsnLeuSerSerValGluAspProGluLysIleLysTrp 579
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
592 .....GGTATTACTGTACAAATGTAGTAAGAAGTGTACGGG 629
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
580 ArgGlnLeuLeuGlyPheArgAsnThrTyrThrCysArgAsnAlaTyrGl 596
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
630 AGGC..... 633
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
596 ySerLysMetLeuValSerSerTyrGluCysIleMetArgAspThrAsnT 613
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
633 ..... 633
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
613 yTrGlnPheCysGluValCysArgLeuGlnGlyPheLysArgMetSerGln 629
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
634 .....AGCTGTACACCAAAAGATGCACATTCATAAAGT 668
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
630 LeuValLysAspValAspLeuTyrValAlaThrProGluValLysGluTyr 646
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
669 AACAGACCTCTATGAAAAAGGATGTGAGTTTGTCTTCCAATCCGCCAGA 718
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
646 rThrGlyAlaTyrSerLysProSerAspPhe.....T 657
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
719 CGGAGAGGCTCTATATGTTTGGCACACATGTT.....GAT 756
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
657 hAspLeuGluThrSerSerTyrTyrAsnTyrThrTyrAsnArgAsnAsp 673

```

```

757 TCTATAGTTGAATTTCTGTACAGACAAACAAACACAAAGAGCTCCAAA 806
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
674 ArgLeuLeuSerGlyAsnSerLysSerArgPheAsn.....ThrAs 687
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
807 CAACCAAAATCAAAAATGCAATCTCCGAAGCACATGGGAAGTGCCTG 856
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
687 nMetAsnGlyLysIleGluLeuArgThrValIleGlnAsnIleSerA 704
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
857 ATTCTGAG.....GACTTTAAGAAACCACTCTCTATGACAACA 894
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
704 sPLysAsnAlaArgGlnLeuLysPheLys..... 713
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
895 CAGCCACCAATCCACCTTCTCATTTGCTGCAGATTGGACAAAGATTGT 944
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
713 ..... 713
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
945 GTGTTTAGTCTCTGACAAATCTGGAAGCATGGCGACTGGTAACCGCTCA 994
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
714 .MetTrpIleLysHisSerAspGlySerValAlaThrAspSerGlyA 730
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
995 ATCGACTGAATCAAGACAGCGCAGCTTTTCTGCTGCAGACAGTTGAGCTG 1044
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
730 snProLeuGln.....ThrValGlnThrPheAspIle 740
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1045 GGTCTCTGGTGGGATGCTCACATTT.....GACAGTGTGCCCA 1085
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
741 ProValTrpAsnAspLysAlaAsnPheTrpProLeuGlyAlaLeuAspHi 757
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1085 TGTACAAAGTCAACTCATACAGATAACAGTGGCAGTGCACAGGACACAC 1135
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
757 sIleLysSerAsp.....PheAsnSerGlyLeuLysSerCysSerL 771
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1135 TCGCCAAAAGATTACCTGCAGCAGCTTCAGAGGAGGACGCTCATCTGAGC 1185
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
771 euLysTyrGlnIleProSerAspAla.....GlnLeuLysSer 783
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1186 GGGCTTCGATCGGCATTTACTGTGATT..... 1212
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
784 GlyAspThrValAlaPheGlnValLeuAspGluAsnGlyAsnValLeuAl 800
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1213 .....AGGAAGAAATATCCAACT..... 1230
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
800 aAspAspAsnThrGluThrGlnArgTyrThrThrValSerIleGlnTyrL 817
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1231 .....GATGGATCTCAAAAT..... 1245
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
817 ySPheGluAspGlySerGluIleProAsnThrAlaGlyGlyThrPheThr 833
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1246 .....GTGCTGCTGACGGA 1259
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
834 ValProTyrGlyThrLysLeuAspLeuThrProAlaLysThrLeuTyrAs 850
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1260 TGGGGAAGACAACTATAAGTGGTGTCTTAACGAGCTCAACAAAGTG 1309
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
850 pTyrGluPheIleLysValAspGlyLeuAsnLysProIleValSerAspG 867
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1310 GTGCCATCATC.....CACACAGTC 1329
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
867 lyThrValValThrTyrTyrTyrLysAsnLysAsnGluGluHisThrHis 883
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1330 GCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAAATGAC 1379
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
884 AsnLeuThrLeuValAlaAla.....LysAlaAlaThrCysThrTh 897
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1380 AGGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGAACAAATGGCTCA 1429
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
897 rAlaGlyAsnSerAlaTyrTyrThrCysAspGlyCysAspLysTrpPheA 914
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1430 TTGATGCTTTTGGGGCCCTTCATCAGGAATGGAGGTCTCTCTCAGCGC 1479
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
914 laAspAlaThrGlySerValGluIleThrAsp.LysThrSerValLysIle 930

```


262 snLeuGlyAspIlePheAlaLysGlyGlyAsnIleAsnValArgAlaAla 278
 1405 GATCAAGTTCAGACAATGGCGCTCATTT.....GATGCTTTTGGGCCCT 1448
 279 Asn...IleArgAsnGlnGlyLysLeuSerAlaAspSerValSerLysAs 294
 1449 TTCATCAGGAATAGAGCTGTCCTCTCAGCGCTCATCCAGCTTGAGAGTA 1498
 294 pLysSerGlyAsnIleValLeuSerAlaLysGluGlyGluAlaGluLeuG 311
 1499 AGGGA...TTAACCTCCAGAACAGCAGCTGGATGAATGCACAGTCATC 1545
 311 yGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGlyLysLeuMet 327
 1546 GFG..... 1548
 328 IleThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLeuSe 344
 1548 1548
 344 rGlyLysGluGlyGlyGluThrThrLeuGlyGlyAspGluArgGlyGluG 361
 1549GACAGCACCGTGGGAAAGGACACT 1572
 361 yLysAsnGlyIleGlnLeuAlaLysLysThrSerLeuGluLysGlySer 377
 1573 TTGTTTCTTATCAGCTGGACAACGAGCCTCCCAATCTTCTCTGG... 1620
 378 ThrIleAsnValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGI 394
 1621GATCCACAGTGGAGAC..... 1635
 394 yAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGlyLysAspIleA 411
 1636 ...AAGCAGGTGGCTTCTAGTCGACAAAACACCAGCAATGGCGTCACTC 1683
 411 IaLysThrGlyGlyPheVal.....GluThrSerGlyHisTrpLeu 424
 1684 CAATATCCAGGCGATGCTTAAGGTGGCACTTGGAAATACAGTCTGCAAGC 1733
 425 SerIleGlyAsnAspAlaAlaValGluAlaLysGluTrpLeuLeuAspPr 441
 1734 AGCTCCAAACCTTGAGC..... 1752
 441 oAspAsnValThrIleSerAsnGlyAsnAspGlnSerGlnLeuLysA 458
 1752 1752
 458 sPAspArgGlyAspSerProAsnLysIleLeuAlaAspAsnLysHisThr 474
 1752 1752
 475 ValAsnAsnLysThrLeuSerThrAlaLeuLysGlyIleGlyValAs 491
 1753CTGACTGTCACTCCGTCGGTCCCA 1777
 491 ntleSerAlaLysLysLysValAsnValThrAlaAspIleAsnValHisA 508
 1778 ATGCTACCTGCCTCCAATTACA..... 1800
 508 snGlyThrLeuThrLeuHisSerGluGlnGlyValGluIleAsnGly 524
 1801 ...GTGACTTCCAAACGACAGGAC...ACCAGCAAAATCCCCAGGCC 1844
 525 AspIleThrSerGluGlnAsnGlyAsnLeuThrIleLysAlaGlySerTr 541
 1845 TCTGGTAGTTTATGCAAAATATTCCGCCAAGGAGCTCCCAATT...CTCA 1891
 541 pValAspValHisLysAsnIleThrIleGlyThrGlyPheLeuAsnIle 558
 1892 GGGCGAGTGTCACGCCCTGATTGAATCAGTGAATGA..... 1929

```
seq_name: sp_bacteria:Q9ZD91
seq_documentation_block:
```


[illegible]

1230	TGATGGTCTCAAATTGTGCCTCAGCGATGGGGAACCAACACTATAA	1279
1231	: : : : : : :	1280
1044	rGlyGlyIleIleThrLeuAspSerAsnGlyAsnAsnLeuThrIles	1061
1280	GTTGGTGCTTTACGAGGTCAAACAAGTGGTGCCATCATCCACACAGTC	1329
1061	erGly.....AsnAsnGly.....Ile	1066
1330	GCTTTGGGCCCTCTGCAGCTCAAGAAGTCAAGCAGCTGTCACAAATGAC	1379
1067	LysLeuGly...SerLysGlyAsnGluLeuSerLeuAsn.....Il	1080
1380	AGGAGGTTTACAGACATATGCTTCAGAT...CAAGTTTCAGAACAAATGCC	1426
1080	eLysGlyLysValThrValThrAsnAspLeuAspIleGlnAsn.....	1094
1427	TCATTGATGCTTTGGGCCCTTTTCATCAGAAATAGGAGTGTCTCTCAC	1476
1095IleHisGlnLeuAsnIleAsnAsnGlyAlaLeu.....	1105
1477	CGCTCCATCCAGCTTGACAGTAAGGATTAAACCCTCCAGAACACCCAGTG	1526
1106PheAspAspGlnSerLeuThrSerAlaLysIleLysAs	1118
1527	GATGAAT...GGCACAGTGATCGTGGACAGCCAGCTGGGAAGACCACTT	1573
1118	nileAsnIleGlyThrValAlaGlyAlaThrTyThrLeuAspAlaI	1135
1574	TGTTCTTATACCTGGACACAGGAGCCTCCCCAAATCCTCTCTGGGAT	1623
1135	le.....	1135
1624	CCCAGTGGACAGACAAGGTGGCTTTGCTAGTGGACAAAAACACAAA...	1671
1136AsnAspAsnPhe.....AspLeuAsnThrSerGI	1145
1672	ATGGCCTACCTCCAATCCCAGGACATGTCTAAGGTTGGCAGTTGGAAAT	1720
1145	yMetValPheLysHisGlnAspSerIleLeuGluLeu.....LysA	1159
1721	ACAGTCTGCAAGCAGCTCACAAACCTTGACCCCTGACTGTACGTCGCCGT	1770
1159	snsrSerAsnThrAsnAspHisThrIleThrLeuThrSerAlaLeuAsp	1175

1330 GCTTTTGGGGCCCTTTCAGAGCTCAAGAACTAGACGAGCTGTCACAAATGAC 1370
1067 LysLeuGly...SerLysGlyAsnGluLeuSerSerLeuAsn.....I1 1080
1380 AGGAGGTTTACAGACATATGCTTCAGAT...CAAGTTCCAGAACAAATGGCC 1426
1080 eLysGlyLysValThrValThrAsnAspLeuAspIleGlnAsn..... 1094
1427 TCATTGATGCTTTTGGGGCCCTTCATCAGAGAAATGAGCTGTCTCTCAG 1476
1095IleHisGlnLeuAsnIleAsnAsnGlyAlaLeu..... 1105
1477 CGCTCCATCCAGCTTGAGAGTAAGGATTAACCTCCAGAACAGCCAGTG 1526
1106PheAspAspGlnSerLeuThrSerAlaLysIleLysAs 1118
1527 GATGAAT...GSCACGAGTATCGTGGACAGCACCGCTGGGAAGACACACTT 1573
1118 nIleAsnIleGlyThrValAlaGlyAlaThrTyThrLeuAspAlaI 1135
1574 TGTTCTTTATCACCTGGACAACGAGCGCTCCCCAAATCCTCTCTGGGAT 1623
1135 le..... 1135
1624 CCCAGTGGACAGCAAGGTGGCTTTTCTAGTGCACAAAACACCAA... 1671
1136AsnAspAsnPhe.....AspLeuAsnThrSerGI 1145
1672 ATGGCCCTACCTCCAAATCCCGAGGCATTGCTTAAGAGTTTGGCACTTGGAAAT 1720
1145 yMetValPheLysHisGlnAspSerIleLeuGluLeu.....LysA 1159
1721 ACAGTCTGCAAGCAGCTCACAAACCTTGACCTGACCTGTCAGTCCCGT 1770
1159 snSerSerAsnThrAsnAspHisThrIleThrLeuThrSerAlaLeuAsp 1175

487 erilethrLeuGlyGlyThrAsnGlyGlyAsnAspAlaLysThrVal 503
553 TATATCC...AATGAGAGATACAGCAGTAAAGATGTTCCAGCAGGTATTAC 599
504 IleAsnLysAspGlyLeuThrIleThrProAlaGlyAsnGlyGlyThrTh 520
600 TGGTCAAAATGTAGTAAAGAGTGTCTCAGGAGGACACCTGTTACACCAAA 649
520 rGlyThrAsnThrIleSerValThrLysAspGlyIleLysAlaGlyAsnL 537
650 GATGCACATTCATATAAGTAACAGGACTC...TATGAAAAGGATGT 693
537 ysaLaillethrAsnValAlaSerGlyLeuArgAlaTyraSpAspAlaAsn 553
694 GAGTTGTTCTCCAAATCCGCCAGCAGAGGAGGCTTCTATAATGTTGC 743
554 PheAspValLeuAsnAsnSerAlaThrAsp...LeuAs 565
744 ACAACATGTTGATTCTATAGTGAATCTGTACAGAACAAACCAACA 793
565 nargHisValGluAspAlaTyrlLysGlyLeuLeu...AsnLeuAsnG 580
794 AAGAAGCTCCAAACAAAGCAAAATCAAAAATGCAATCTCCGAAGCACATGG 843
580 luLysAsnAlaAsnLysGln... 586
844 GAAGTATCCGGTATCTCAGGACTTTAAGAAAACCACTCTATGACAAC 893
587ProLeuValTh 590
894 ACAGCACCACAAATCCCACTTCTCATTGCTGCAGATGGACAAGAATTC 943
590 rAspSerThrAlaAlaThrValGlyAspLeu...ArgLysLeuG 604
944 TGTGTTTAGTCTTGAATCTGGAAGCATGCGACTGTGTACCCGCTC 993
604 lyTrpValSerThrLysAsnGly...ThrLysGluGluSer 617
994 AATCGACTGAATCAACAGCAGCCAGCTTTCTCCTGCTGCAGACATTTAGCT 1043
618 AsnGlnValLysGlnAlaAspGluVal...Le 627
1044 GGGCTCTGGTGGGATGGTACATTTGCAGCTGTCGCCATGTACAAA 1093
627 uPheThrGlyAlaGlyAlaAlaThrValThrSerLysSerGluAsnGlyL 644
1094 GTGAACATCATACAGATAAACAGTGGC... 1119
644 yshisThrIleThrValSerValAlaGluThrLysAlaAspCysGlyLeu 660
1120 ...AGTCAGAGGACACACTCGCCCAAGATTACCT... 1152
661 GluLysAspGlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAs 677
1153GCAGCAGCTTCAGGAGGAGTCCATCTCGCAGCGGCTC 1192
677 pAsnValLeuThrValGlyAsnAsnGlyThrAlaValThrLysGly... 692
1193 GATCGGCATTTACTGTAGTAGGAAGAATATCCAACTGATGATCTGAA 1242
693GlyPheGluThrVal...LysThrGlyAlaThrAspAlaAspArg 706
1243 ATGTGCTCTGCGGATGGGAGACACACTATAAGTGGTCTTTAA 1292
707GlyLysValThrValLysAspAlaThrAlaAs 717
1293 CGAGGTCAAAAGTGGTCCCATCATCCACAGCTCGCTTTGGGCCCT 1342
717 nAspAlaAspLysValAlaThrValLysAspValAlaThrAlaIle 734
1343 CTCGACTCAAGAACTAGAGAGCTGTCCAAAATGACAGAGGTTTACAG 1392
734 snSerAlaAlaThrPheValLysThrGluAsnLeuThrThrSerIleAsp 750

1393 ACATATGCTTCAGATCAAGTTCAGAACAAATGGCCTCATTCATGCTTTTGG 1442
751GluAspAsnProThrAspAsnGlyLysAspAsp... 761
1443 GGCCTTTTCATCAGAAAT...GGAGCTGTCTCTCAGCGCT 1480
762 AlaLeuLysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuL 778
1481 CCATCCAGCTTGAGAGTAAGGATTATCCCTC...CAGAACAGC 1521
778 ysaLysArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeu 794
1522 CAGTGAATGAATGGCAGATGATCGTGGACAGCACCCTGGGAAAGGACAC 1571
795 GluValLysThrAlaLysValSerAspThrLeuThrIleGlyLysnTh 811
1573 TTTGTTTCTTATCACCTGGCAACAGCCTCCCAATC... 1611
811 r...ProThrGlyGlyThrThrAlaThrProLysValAsnIleThrS 826
1611 1611
826 erThrAlaAspGlyLeuAsnPheAlaLysGluThrAlaAspAlaSerGly 842
1612CTTCTCTGGGATCCAG 1628
843 SerLysAsnValTyrLeuLysGlyIleAlaThrThrLeuThrGluProSe 859
1629 TGGACAGAACCAAGTGGCTTTGTA...GTGGACAAAACACCA 1669
859 rAlaGlyAlaLysSerSerHisValAspLeuAsnValAspAlaThrLysL 876
1670 AATGGCCTTACCTCCAAATCCAGGCAATGCTAAGTTGGCAGCTTGGAAA 1719
876 ySerAsnAlaAlaSerIleGluAspValLeuArgAlaGly... 889
1720 TACAGCTGCAAGCAAGCTCACAAACCTTACCCTGACTGTCAGCTCCCG 1769
890 TrpAsnIleGlnGlyAsnGlyAsnValAspTyrValAlaThrTyrAs 906
1770 TGGCTCCAATGCTACC...CTGCCCTCCAATTACAGTGA 1804
906 pThrValAsnPheThrAspSerThrGlyThrThrThrValThrValT 923
1805 CTTCAAAACGAACAGGACACCAAGCAAAATCCCAAGCTCTGGTAGTT 1854
923 hrGlnLysAlaAspGlyLysGly... 930
1855 TATCCAAATATTCGCCAAGGAGCTCCCAATTCACAGGCCAGCTGCAC 1904
931AlaAspValLysIleGlyAlaLys...Th 939
1905 AGCCTGATTGAATCAGTGAATGGAACACAGTTACC...TTGGAACATAC 1951
939 rSerValIleLysAspHisAsnGlyLysLeuPheThrGlyLysAspLeuL 956
1952 TGGATATGGAGCAGGTGCTGATGCTACTAAGGATGACGGT... 1992
956 ysaAlaAsnAsnGlyAlaThrValSerGluAspAspGlyLysAspThr 972
1993GTCTACTCAAGTATTTCACAACCTTATGCACAGCAATGG 2030
973 GlyThrGlyLeuValThrAlaLys...ThrValIleAspAlaValas 987
2031 TAGATACAGTGAATAAGTGGGCTCTGGGA... 2061
987 nLysSerGlyTrpArgValThrGlyGluGlyAlaThrAlaGluThrGlyA 1004
2062GGAGTTAAGCAGCAGCAGGAGAGTGTATACCCAGCAGAGTGA 2106
1004 lathrAlaValAsnAlaGlyAsnAlaGluThrValThrSerGlyThrSer 1020

```
2107 GCAGTGTACATACCTGCTGCTGATTGAGAATGATGAATACAAATGAATCC 2156
    |||
    |||
1021 ValAsnPhelysAsngly.....AsnAl 1028
2157 ACCAAGACCTGAATTAATTAAGGAT.....G 2182
    |||
    |||
1028 aThrThrAlaThrValSerLysAspAsnglyAsnIleAsnValLysTyrA 1045
2183 ATGTT.....CAACACAAGAGTGTGT 2205
    |||
    |||
1045 spValAsnValglyAspGlyLeuLysIleGlyAspAspLysIleVal 1061
2206 TTCACAGACATCC.....TCGGCAGGCTCATTTGTG..... 2238
    |||
    |||
1062 AlaAspThrThrLeuThrValThrGlyGlyLysValSerValProAl 1078
2239 ....GCTTCTGATGCCAATGCTCCCATACCTGATCTCTCCACCTG 2284
    |||
    |||
1078 aGlyAlaAsnSerValAsn..... 1085
2285 GCCAATCACCACCTGAGCGGGAATTCAGGGGGCAGTCTCATTAT 2334
    |||
    |||
1086 ..AsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsnAsn 1101
2335 CTGACTTGACAGCTCCTGGGATGATTATGACCATGGAACAGCTCAAA 2384
    |||
    |||
1102 LeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyGluSerGluG1 1118
2385 GTATATC..... 2391
1118 yGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysAlaG 1135
2392 .....ATTGAAATAGTACAAGTATTCTTGATCTCAGACAAAGTTC 2433
    |||
    |||
1135 LysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSerLeu 1151
2434 AATGAATCTCTCAAGTAACTACTGCTCTCATCCCAAGAGCCAA 2483
    |||
    |||
1152 GlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAlaAs 1168
2484 C.....TCTGAGGAAGTCTTTTGTGTTTAAACACAGAAACATTTCTTT 2526
    |||
    |||
1168 nGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrIleT 1185
2527 .....GAAATGCCACAGAT 2541
    |||
    |||
1185 hrLeuAlaAsnGlyAlaAlaGlyThrAspAlaSerAsnGlyAsnThr 1201
2542 CTTTTTCATTCCTATTCAGGCTGTTGTAA..GTCGATCTCAAAATCAGAAA 2590
    |||
    |||
1202 IleSerValThrLysAspGlyIleSerAlaGlyAsnLysGluIleThrAs 1218
2591 TATCCACATTCGACGAGTATCTTTGTTTATCTCCTCACAGACTCCGCCA 2640
    |||
    |||
1218 nValLys.....SerAlaLeuLysThrTyrLysAspThrGlnAsnThrA 1233
2641 GAGACACTAGTCTGTATGAACGCTGCTCTGCTCTTATATTCATAT 2690
    |||
    |||
1233 laAspGlu..... 1235
2691 CAACAGCACCATTCTCGCATTCACATTTTAAAAATTTATCTGGAAGTGA 2740
    |||
    |||
1236 ThrGlnAspLysGluPheHisAlaAlaValLysAsnAlaAsnGluValG1 1252
2741 T 2741
1252 u 1252
seq_name: sp_invertebrate:Q9U7P4
seq_documentation_block:
ID Q9U7P4 PRELIMINARY;
AC Q9U7P4; PRT; 494 AA.
```

```
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 55.0 KDA PROTEIN (FRAGMENT).
OS Eufolliculina uhligi.
OC Eufolliculina.
OC NCBI_TaxID=46026;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99444925; PubMed=10517337;
RT Markmann-Mullisch U., Reiss B., Mullisch M.;
RT "Cell type-specific gene expression in the cell cycle of the dimorphic
RT ciliate Eufolliculina uhligi.";
RL Mol. Gen. Genet. 262:390-399(1999).
DR EMBL; AF121336; AAF13350.1; -.
DR InterPro; IPR002035; VWFA.
DR InterPro; IPR003015; HLH_Myc.
DR Pfam; PF00092; vwa; 1.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
DR PROSITE; PS00334; VWFA; 1.
DR SMART; SM00327; VWA; 1.
KW Hypothetical protein.
FT NOW_TER 1
SQ SEQUENCE 494 AA; 54960 MW; 4A69C8B8892D7B24 CRC64;
```

```
alignment_scores:
  Quality: 140.50      Length: 420
  Ratio: 0.614        Gaps: 19
  Percent Similarity: 54.524    Percent Identity: 22.381
```

alignment_block:

US-09-049-696-20 x Q9U7P4 ..

align seg 1/1 to: Q9U7P4 from: 1 to: 494

```
895 CAGCCACCAAAATCCACCTCTCTCATTCGTCGAGATTGGACAAAGATTGT 944
    |||
    |||
76 GluSerProAlaGlnThrSerGluAlaSerArgSerGlyValAspIleVa 92
945 GTGTTTACTCTTGCACAAATCTGGAACCATGCCAGCTGGTAACCGCTCA 994
    |||
    |||
92 1Cys...ValIleAspValSerGlySerMet...GlnGlyGluLysIle. 106
995 ATCGACTGAATCAAGCCAGCGCTTTCTCTGCTGCAGACAGTTGAGCTG 1044
    |||
    |||
107 ..GlnLeuValGlnThrThrLeuAsnPheMetValGluArgLeuSerPro 122
1045 GGGTCTCTGGTGGATGGTGACATTTCACAGTGTGCCCATGTACAAAG 1094
    |||
    |||
123 AlaAspArgIleCysLeuIleSerPheSerAsnAspAlaThrLysIleSe 139
1095 TCAACTCATACAGATAAACAGTGGCAGTGACAGGACACACTCGCCAAAA 1144
    |||
    |||
139 rArgLeuValGlnMet...SerProLysGlyLysLysGlnLeuLysSerM 155
1145 GATTACCT...GCAGCAGCTTCAGGAGGAGCTCCATCTGACGGGGCTT 1191
    |||
    |||
155 etIleProArgLeuValAlaSerGlyGlyThrAsnIleValGlyGlyLeu 171
1192 CGATCGGCATTTACTGTATTAGGAAGAAATATCCAACTGATGGA...TC 1238
    |||
    |||
172 GluTyrGlyLeuGlnAlaLeuArgGlnArgThrIleAsnGlnLeuSe 188
1239 TGAATTTGCTGCTCACGGATGGGGAAGACAACTATACTGGGTGCT 1288
    |||
    |||
188 rSerIleIleLeuLeuSerAspGlyGlnAspAsnGlyThrThrValL 205
1289 TTAACAGAGTCAAA.....CAAAGTGTGCC 1314
    |||
    |||
205 euGlnArgAlaLysAlaThrMetAspSerIleValIleArgAspAspTyr 221
```

```
alignment_block:
US-09-049-696-20 x Q9V3J6 ..
```

```

    alignment_block:
      US-09+049-696-20 x Q9V3J6 ..
      Percent Similarity: 46.166
      Ratio: 0.496
      Quality: 140.50
      Length: 6135
      Percent Identity: 21.533
  
```

Tue Apr 2 09:40:00 2002

Align seq 1/1 to: Q9V3J6 from: 1 to: 721

1210 ATTAGGAAGAATAATCCA.....ACTGATGACTCAATTTGTGCTGCT 1253
148 ValLysLysArgPheProGlyLeuAsnAsnLeuValGluAlaGlyLeuLeu 164
1254 GACGGATGGGGAAGACAACACTATAAAGTGGGTGCTTTAAACGAGTCC... 1299
164 uasnAspAsnGlu...LysThrIleIleGluThrMetAsnLysAlaPheP 180
1300AAACAAGTGGTGCATCATCCACACAGTCGCTTTGGGG 1338
180 roArgProSerLysHisTrpLeuProIleValTrpAlaAlaSerIleLe 196
1339 CCCTCTCGAGCTCAAGAA..... 1356
197 ThrArgAlaArgLysGluGlyArgIleArgAspAspPheAlaValLysTh 213
1357CTAGAGGAGCTCTCCAAATGACAGGAGTTTACAGACATATGCTT 1402
213 rIleIleAspGluLeuAsnLysPheArgGly..... 223
1403 CAGATCAAGTTCAGAACATGGCCCTCATGTATGCTTTTGGGGCCCTTCA 1452
224GlnCysGlyLeuLeuIleSerTyrAspThrIleSer 235
1453 TCAGGAATGGAGCTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGG 1502
236 ValProLeuValTyrThrGlnValValThrLeuAlaValTyrSerTyrPh 252
1503 ATTAACCC...CTCCAGAACAGCCAGTGGATGAATGGCAGCATGTCGTGG 1549
252 eLeuThrCysCysMetGlyGlnGlnTrpThrAspGlyLysValValGlyA 269
1550 ACAGCACCC...GTGGGAAGGACACTTTGTTCTTATCACTCGGACGACG 1596
269 snThrThrTyrLeuAsnLysValAspLeuTyrPheProValPheThrThr 285
1597 CAGCGTCCCAAACTCTTCTCTGG..... 1620
286 Leu.....GlnPhePhePheTyrMetGlyTrpLeuLysValAlaGluSe 300
1621GATCCAGTGGACAGCAAGCAAGTGCGCTTT.....G 1651
300 rLeuIleAsnProPheGlyGluAspAspAspPheGluValAsnTrpM 317
1652 TAGTGGCAAAAACACCAAAATGGCTTACCTCCAAATC..... 1689
317 etValAspArgAsnLeuGlnValSerTyrLeuIleValAspGluMetHis 333
1690CCAGGCATGTCTAAGTGTGGCACTTGG..... 1716
334 HisAspHisProGluLeuLysAspGlnTyrTrpAspGluValPhePr 350
1717AAATACAGTCTGCAAGCA..... 1734
350 oAsnGluLeuProTyrThrIleAlaAlaGluArgPheArgGluAsnHisP 367
1735 ...AGGTCAACAACCTTGACCTGACTGTACAGTCCCGTGGTGCCTCAATGCT 1782
367 roGluProSerThrAlaLysIleGluValProLysAsnAlaIa..... 381
1783 ACCCTGCGCTCCAATTTACAGTCACTTCCAAAACGAAACGACGACACAGCAA 1832
382MetProSerThrMetSerSerValArgIleAspGluMetAlaAs 396
1833 ATTCCCCAGCCCTCTGGTAGTTATTGCA.....AATATTCGCCAAG 1873
396 pAspAlaSerGlyIleHisPheSerAlaGlyLysnGlyLysMetArgLeuA 413
1874 GAGCCCTCCCAATCTTCAGGGCGCAGTGTACAGGCCCTGATTGAATCATGT 1923

[illegible]

795	GlyGlnAspAlaAlaTrpTyProGlyAspAsnTyValAspIleIleG1	809
1968	TGCTGATGCTACAAAGGATGACGGTGCTCTACTCAAGGTATTTCACAAC	2017
809	yGluAspIleTyrgluGluLysAlaGlnTySerProTyAlaAspA	826
2018	ATGACACGAATGGTGTAGATACAGTCTAAAGATGCGCGCTCTGGGAGAG	2067
826	he.....AlaLysAlaLeuLysTyThr	833
2068	AACGACGCCAGACGAGAGTGATACCCAGCAGAGTGGCAGCACTGTACAT	2117
834	Asn...AlaArgLysMetIleAlaLeuThrGluCysGlyTyrIleProAs	849
2118	ACCTGGCTGGATTCAGAGATGATGAATCAATGGGAAT	2154
849	pProAlaValLeuLysGlnGluGlyIleSerTrpSer	861
seq_name: sp_bacteria:09PHN8		
seq_documentation_block:		
ID	Q9PHN8	PRELIMINARY; PRT; 978 AA.
AC	Q9PHN8;	
DT	01-OCT-2000 (TrEMBLrel. 15, Created)	
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)	
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)	
GN	POSSIBLE LIPOPROTEIN.	
OS	Campylobacter jejuni.	
OC	Bacteria; Proteobacteria; epsilon subdivision; Campylobacte	
OX	NCBI_TaxID=197;	
RN	[1]	
RC	SEQUENCE FROM N.A.	
RC	STRAIN=NCCT 11168;	
RC	MEDLINE=20150912; PubMed=10698204;	
RA	Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C	
RA	Basham D., Chillingworth T., Davies R.M., Feltwell T., Hoir	
RA	Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.	
RA	Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.	
RA	Whitehead S., Barrell B.G.;	
RT	"The genome sequence of the food-borne pathogen Campylobact	
RL	reveals hypervariable sequences."	
RL	Nature 403:665-668(2000).	
DR	EMBL; AL139075; CAB75265.1;	
KW	Complete proteome.	
CC	SEQUENCE 978 AA: 102498 MW: 06DAD1A5FA433609 CRC64:	

```

alignment_scores:
  Quality: 140.00
  Ratio: 0.328
  Percent Similarity: 42.277
  Length: 1010
  Gaps: 46
  Percent Identity: 19.010

alignment_block:
  US-09-049-696-20.x Q9PHN8 ..
  Align seg 1/1 to: Q9PHN8 from: 1 to: 978
  7 ACAGGGAGATGTACAGCAATGGGCCATTTAAGAGTTCTGTGTTTCATCTT 56
  |||||::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
  26 ThrGlyGluSerIleTyrPheGlyAsnAlaLysIleSerSerPheValAs 42
  ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
  57 G.....ATTCCTTCACCTTCTAGAGGGGCCCTCAGTAAT..... 90
  ::: ::: ::: ||||| ||| |||
  42 nSerGlyThrIleLysSerLysGlnGlyAlaGlyValAsnIleSerGlnG 59
  ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
  91 .....TCACCTCATTCCTGAACAACAATGCTGTATGAAGCGATTCTCGTT 135
  ||||: ::::: ||||| ||||| ||||| ||||| ||||| |||||
  59 lyThrSerIleGluAsnPheAsnThrGly...ThrGlyIleIleGlu 74
  ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
  136 GCAATCGACCCCAATGTGCCAGAGACATGAACACTCATTTCAACAATAA 185

```


75 GlyLysArgMetGlyValAsnValArgSerThrIleAsnThrPheValAs 91
186 GGACATGGTGACCCAGCATCTCTGTATCTGTTTGAAGCTACAGAAAGC 235
91 nAsp.....GlyLeuIleAlaAlaThrAsnAspG 101
236 GATTATTATTCAAA...AATGTTGCCAATTTTGTATCTCTGAA...ACATGG 279
101 lyleGlnIleAsnAlaAsnValLysThrLeuIleAsnLysGlyThrIle 117
280 AAGACAAAGGCTGACTATGTGAGACCA.....AACTTGAGACCTA 320
118 LysGlyAspAlaIleSerIleArgSerLeuGlyGlyThrIleGluThrLe 134
321 CAAAAATGCTCATCTTCTGTTGCTGAGTCTACTCTCTCCAGGTAATGAT 370
134 uThrAsnGluGlyIleMetThrGlyLysSerAlaGlyIleThrMetAsnA 151
371 AACCTACATGACGACATGGGCAACTGTGGAGAGAGGTGAAGATC 420
151 rgSerLeuValLysThrLeuThrAsnSerGlyThrIleAsnGlnAsn 167
421 CACCTACTCTGATTTTCATTGTCAGGAAAAAGTTAGCTGAATATGACC 470
168 SerAlaThr.....TrpSerAlaGlyIleLysLeu..... 177
471 ACAAGGTAGGGCATTTCTCCATGAGTGGGCTCATCTACGATGGGAGTAT 520
177 177
521 TTGACGAGTACATATGATGAGAAATCTACTATCCATCCATGGAGAATA 570
178GluAsnGlySerIleIleGluAsnIleIleAsnThrGlySerIle 192
571 CAAGCA.....GTAAGATGTTACAGAGGTATTACTGTACAAA 608
193 ArgSerAsnAlaPheGlyIleIleSerValThrGlyGlyLysPheGlyThrLe 209
609 TGAGTAAAGAGTGTGACGGAGGACGCTGTTACACCAA..... 648
209 uThrIleLys.....AspGlyGlyMetValThrGlyLysThrSerAlaI 224
649AGA 651
224 leGlyValGlyArgSerGlnThrLeuGlyAspLeuThrIleAspGlyArg 240
652 TGACATTCATTAAGTAAGTAACAGGACTCTATGAAAGGATGTGAGTTGT 701
241 SerAsnAsnGlyThrValSerGlyIleThrSerGluGluHisGlyIleLe 257
702 TCTCCATCCCGCCAGACGAGGAGGCTTCTATTAATGTTTGCACAAATG 751
257 uLeuGluAsn..... 260
752 TTGATCTATAGTTGATTTCTGTACAGAACAAACACACAAAGAGCT 801
260 260
802 CCAACAAAGCAAAATCAAAATGCAATCTCCGAGACATGGGAAGTGT 851
261 ...AsnSerArgThrGlnLysIleGluLeuLysAsn...GlyGlyIleII 275
852 CCGTGTATCTGAGACCTTTAAGAAACCACTCTATGACAAACACAGCCAC 901
275 eLysGlyAsnIleAsp..... 280
902 CAATCCACCTTCTCTATCTGTCAGATTCGACAAAGAAATTTGTGT... 948
281GlyIleArgLeuIleAsnSer 287
949TTAGTCTTGACAAATCTGAGAGCATGGCGAC 980
..... 999

289 AlaSerLeuSerGlyGluMetIleLeuSerGlyGluGlySerArgValGI 304
981 TGGTAACCGC...CTCAATCGACTCAATCAAGCAGCGCCAGCTTTTCTCTGC 1027
304 uGlyGlyArgGlyValGlyIleLeuAsnArgSerGlyLysIleGluGlys 321
1028 TGCAGACAGTTGAGCTGGGCTCTCTGGGTT..... 1056
321 erIleLysValGluAspGlyAlaThrValThrAlaThrSerAsnArgAla 337
1057GGGATGGTGACA..... 1068
333 IleAlaAsnSerGlySerGlySerIleThrGlyIleThrValSerGI 354
1068 1068
354 yLysAsnThrLysLeuGluGlyAsnIleIleAsnThrGlyAsnAlaSerI 371
1069TTTGACAGTGTGCCCATGTACAAAGTGAATC 1101
371 leGlySerAspIleLysIleGluGlyAlaLysValGluGlyGlyLeu 387
1102ATACAGATAAACAGTGG 1118
388 ValAsnGlnGlyAsnGlySerIleSerGlySerValGlnValSerGlyGI 404
1119 CAGTGACAGGACACACTCGCCAAAGATTACCTGCAGCAGCTTCAGGAG 1168
404 ySerIleAspSerIleThrAsnGluGlyAsnGlyAlaIleSerGlyS 421
1169 GGAGCTCCATCTGC.....AGCGGGCTTCGATCGGCATTTACTGTGATT 1212
421 erIleThrValThrLysAspSerLysLeuAspSerIleThrAsnThrSer 437
1213 AGGAAGAAATATCAACTGATGATCTGAAATTTGCTGCTGACGAGTGG 1262
438 ThrSerThrGlyIleSerGlySer.....IleThrAsnAs 450
1263 GGAAGACAACTATAAGTGGTGTCTTTACGAGGTCAAAACAAAGTGTGT 1312
450 nSerAspAsnLysLeu.....GluIleSerAsnSerGlyA 462
1313 CCATCATCCACACAGCTCGCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAG 1362
462 snIle..... 463
1363 GAGCTGTCCAAATGACAGAGGTTTA.....CAGACATATGCTTTCAGA 1406
464GlyGlyLysIleGluSerThrGlySerAlaAs 474
1407 TCAAGTT.....CAGAAATGGCCTCATTCATGCTTTTGGGGCCC 1447
474 pMetValIleSerAsnSerAsnGlyGlyThrIleSerGly...GlyIleS 490
1448 TTTTCATCAGGAAATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAT 1497
490 erSerSerGlySerGlySerThrSerIleSerAsn.....Ser 502
1498 AAGGGATTAAACCTCCAGAACACAGCAGTGGATGAATGGCAGCTGATCGT 1547
503 GlnGlySerThrIleAsnAsnGlyIleThrValSerGlySerAlaGlnVa 519
1548 GGAC.....AGCACCGTGGGAAGGACACACTTTGTTCTTATCA 1585
519 lGluIleSerAsnGlnGlySerValGlyLysAsp..... 530
1586 CCTGGACAACGACGCTCCCAAAATCTTCTCTGGGATCCCGAGTGGACAG 1635
530 530
1636 AAGCAAGTGGCTTTGTAGTGACAAAAACACAAAAATGGCTTACCTCCA 1685
531 ...GluAsnGlyAsnThrValThrAsnAsn..... 539

```
1686 AATCCAGGCAATGCTAGGTTGGC.....ACTTGGAAATACAGCTGCG 1729
|||||
540 .....GlySerGlySerValGlyIleLeuYAspThrLeuValSerThrA 554
1730 AAGCAAGCTCAACAACCTTACGACCTGACTGTGCAG.....TCCGCTGCG 1773
|||
554 spLysAsnThrGlyLysLeuAsnThrValIleGlyGlySerArgAla 570
1774 TCCAAATGCTACCGCTGCTCCATTCACATGAGTCTTCCAAACCAACAAGGA 1823
|||||
571 PheAsnValLysValGluAsnIleThrVal...AspGlnSerAsnValAs 586
1824 CACCAGCAAAATCCCCAGCCCTCTGGTAGTATTGCAAAATATTGCGCCAG 1873
586 pLeuGluGluLeu..... 590
1874 GAGCCTCCCAATTTCTACGGCCGAGTGTACAGCCCTGTATTAATCAGTGG 1923
591 .....AsnAspIleAsnAsnIleIleSerGlyVal 600
1924 AATGGAAACACATTACCTTGGAACTACTGGATAATGAGCAGGCTGCTGA 1973
601 AsnGluAsnAsnIleGly...AsnIleGlyThrAsnGlySerGlyGluI 616
1974 TGCTACTAAGGATGACGGGTGTCTACTCAAGGTATTTTCAACAACCTTATGACA 2023
616 eSerLeuSerPheAspProIleThrGlyLysLeuThrThrAspPheAsnL 633
2024 CGAATGTAGATACAGT...GTAAAGTGGGGCTCTGGGAGGAGTTAAC 2070
633 euAsnAlaSerIleSerGlyAlaThrPheArgSerLeuIleSerThrThr 649
2071 GCAGCCAGCGGAGGTGATACCCAGCAGAGTGGAGCACTGTACATACC 2120
650 SerArgSerThrPheIle..... 656
2121 TGGCTGATTGAGATGAAATACATACATGGAATCCACCAAGACCTGAMA 2170
657 .....AspAsnValMetGlyAsn..... 662
2171 TTAATAGGATGTTGTACACAGCAAGTGTCTTACAGCAAGACATCC 2220
663 .....SerMetGlnSerPheAlaLeuAlaSerSerSer 673
2221 TCGGGAGGCTCAATTTGGTCTCTGATGTCCCAAT.....GCTCCCAT 2264
674 LysSerGlnSerIleAlaMetSerGluLysGlyAsnLeuThrAlaAspAl 690
2265 ACCTGATCTCTCCCACTGGCCAAATACCCGACCTGAAGCGGAAATTC 2314
690 aSerAspTyr.....IleLysSerAspLeuA 699
2315 ACGGGGCGAGTCTCATTAATCTGACTTGGAGCTCCTGGGATGATTAT 2364
699 snAsnGlySer.....Tyr 703
2365 GACCATGGAACGCTCACAAGTATATCATTCGAATAGTACAAGTATTCT 2414
704 GlySerAsnLysGluHisSerLeuPheIleLeuProTyrThrSerSerGl 720
2415 TGATCTCAGACAGCAAGTTCAATGAATCTCTTCAAGTGAATACTACTGCTC 2464
720 nasnValGluLeuSerLeuAsnGluSerLysGlyHisThrLysGlyT 737
2465 TCATC.....CCAAAGGAGCCAACTCTGAGGAAGTCTTT 2499
737 hrIleIleGlyTyrSerThrLeuLysAspSerGlyIleTyrGlyValTyr 753
2500 .....TTGTTTAAACCAAGAAA 2516
754 AlaGlyTyrGluAspThrLysMetGlySerThrThrPheAspIleAsnAs 770
```

```
2517 CATTAACCTTT.....G 2527
|||||
770 nArgThrTyrTyrAlaGlyLeuLysTyrPheAsnThrLeuPheThrThrG 787
2528 AAATGCGACAGATCTTTTCATTCCTGCTATTTCAG.....GCTGTTGAT 2568
|||||
787 LuLysGlyGlnGluValThrIleLysAlaGlnGlyLysAlaAlaLeuIle 803
2569 AAGTCTGATCTGAAATCAGAAATATCCAAAC 2598
|||||
804 LysAsnAspLeuThrGluLysIleGlyAsn 813
seq_name:..sp_bacteria:052781

seq_documentation_block:
ID 052781 PRELIMINARY; PRT: 1112 AA.
AC 052781;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-1998 (TrEMBLrel. 06, Last annotation update)
DE SURFACE LAYER PROTEIN.
GN SAMP2.
OS Campylobacter fetus.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=196;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CIP 5396T;
RA Guesdon J.L., Casademont I., Chevrier D.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF048699; AAC04580.1;
SQ SEQUENCE 1112 AA; 112503 MW; DA9D9C2F62602091 CRC64;

alignment_scores:
Quality: 139.50 Length: 740
Ratio: 0.407 Gaps: 32
Percent Similarity: 46.351 Percent Identity: 20.270

alignment_block:
US-09-049-696-20 x 052781 ..
Align seg 1/1 to: 052781 from: 1 to: 1112
133 GTTGAATCGACCCCAATGTGCCAGAGATGAACACATCTTCATCAACAAT 182
|||||
466 ValThrValAspGlyAsnThr...LysAspLeuThrLeuSerLeuLysLy 481
183 AAAGGACATGGTGACCCAGCATCTCTGTATCTGTTTGAAGCTACAGGA 232
|||||
481 sAlaGlnSerValThrGluSerSer..... 489
233 AGCGATTTTATTTCAAAAATGTTGCCATTTTGATTCCTGAAACATGGAAG 282
|||||
490 .....PheLysAsnIleAlaLysThrValThrGluSer..... 500
283 ACAAGGCTGACTATGTGAGACCAAACTTGACACCTACAAAAATGCTGA 332
|||||
501 .....AsnGlyAsnValGluThr.....ValAs 508
333 TGTCTGTTGTGAGTCTACTCTCCAGGT.....AATGATGAAC 373
|||||
508 nIleLeuAlaAsnAsnAlaThrAspLysAlaValThrIleAsnAspGlu 525
374 CCTAC...ACTGAGCAGATGGGCAACTGTGGAGAGAAGGTGAAAGGATC 420
|||||
525 erLeuLysThrIleAsnPheSerAspValAspLysGlyAlaSerVal 541
421 CACCTCACTCTGATTTTCATTCATTCAGGAAAAAAGTTAGCTGAATATGACC 470
|||||
542 ArgAlaLysGlyLysIleValAlaAspLysAla..... 552
```

471 ACNAGGTAGGGCATTTGTCCATCAGTGGGCTCATCTACGATGGGAGTAT 520
552
521 TTCACAGGTACAATAATGATGAGAAATCTACTTATCCAAATGGAAGAATA 570
553
571 CAAGCAGTAAAGATGTTTCACGAGGTATTACTGTTACAAATGTTAGTAAGAA 620
563 ThrAlaAla...AlaAspAlaValGlnAlaAlaAsnAlaThrLysI 578
621 GTGTCAGGAGGAGCGTGTACACAAAAGATGCACATTCAATAAAGTAA 670
578 eAspIleAsnAlaAlaLysAspThrValGlyLeuThrLeuGlyVala 595
671 CAGGACTCTATGAA.....AAAGATGTGAGTTGTCTC 705
595 laLysLeuThrAspLeuThrValAsnAsnLysGlyAlaPheAlaLeuThr 611
706 CAATCCCGCCACAGGAGAGGCTTCTATA.....ATGTTGC 743
612 GlyAlaAsnAlaThrAspLeuAspSerValLysAsnLeuSerValAsnTh 628
744 ACAACATGTTGATTCTATAGTTGAATTCGTACAGAACAAACCAACA 793
628 rGluGlyAlaPheSerIleAlaThrSerLeuLysAsnLeuAsnA 645
794 AAGAAGCTCCAAACAGCAAAATCAAAAATGCAAT.....828
645 snLeuSerLeuAsnGlyValSerAlaAspLeuAsnSerValAsnValGly 661
829CTCGAAGCACATGGGAAGTATCCGTGATCTTGAGCACTT 869
662 ThrAlaThrLeuAlaSerLeuGluAlaAsnIleAsnValSerGlyGluPh 678
870 TAAGAAACCACTCCATCATCAACACAGCACGCCCAAAATCCACTCTCAT 919
678 elysLeuGlyThr.....ThrThrAlaLysGlyAspValaspPheAsnI 593
920 TGCTGCAGATTGGACAAGAATTTGTTAGTCTTCGACAAATCTGA 969
693 leGluAsnValGlyAlaLeuThrLeuGlyAlaIleThrSerSerThrGly 709
970AGCATGGCGACTGGTAAC.....987
710 AsnAlaSerValIleIleSerSerAlaThrGlyAsnValThrLeuGlyAl 726
988CGCTCAATCGACTGAATCAAGCAG 1012
726 aValSerAlaThrGlnGlyAsnLeuThrLeuAsnAlaGlyAsnThrLeuG 743
1013 GCCAGCTTTTCCTG.....CTGCAGACAGTTGAGCTG 1044
743 lYAsnIleThrIleGlyAlaLeuAlaGlyAspIleValSerValAspLeu 759
1045 GGCTCTGGTGGGATGTCATTTGACAGTGTGCTGCCCATCTACAA... 1092
760 GlyGlyValLeuGlyThrIleAsnSerAlaSerGlyAsnLysValGluI 776
1093AGTGAACATACAGATAACAGTGGCAGTGACAGGACACAC 1135
776 eThrSerAsnGluValThrTyrValGlySerGluIleSerLysAsnValV 793
1136 TCGCCAAAAGATTACTCGACAGCTTCAGGAGGAGCTCCATCTGCAAC 1185
793 aIcLu.....IleThrAlaAlaAlaGlyGlyThrAspLeuAsnAla 806
1186 GGCCTTCGATCGGCATTTACTGTGATTAGGAAGAATAATCCAACTGATG 1235
807GlnValIleGlyAlaAlaAlaAspAl 817
1236 ATCTGAATTTGCTGCTCAGCGATGGGAGACACACACTATATAGTGGGT 1285

817 aLeuThrIleIleGlyLysGlyAspThrGlnThrIleThrAlaSerGly. 833
1286 GCTTTAACGAGGTCAAAACAAAGTGTGCCATCATCCACAGTCCGCTTTG 1335
834AspLeuSerGlyGlyThrLeu...ThrLeuThrLeu 844
1336 GGGCCCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAAATGACAGGAG 1385
845 ThrAspAlaThrLysLeuSerSerLeuAspIleSerGlyValLysGlyLe 861
1386 TTTACAGACATATGCTTCAGAT.....CAAGTTTCAGAACAAATG 1423
862 uSerAlaAlaThrAlaIleAspLeuLysAsnValSerValGluAsnLysL 878
1424 GCTCATGATGCTTTTGGGCCCTTTCATCAGGAAATGGAGCTGTCTCT 1473
878 euIleValAspIleGlnGly.....SerAspAlaAlaGluThrIleThr 892
1474 CAGCGCTCCATCCAGCTTGAGAGTAGGATTAACCTCCAGNACAGCCA 1523
893 AlaAsnSerThrSerAlaThrLeuThrAlaIleThrLeuSerGlyAspLe 909
1524 GTGGATGAATGGCACAGTGCATCGTGACAGCACCGGTGGAAAGGACACTT 1573
909 uGlyGlyGlyAlaAsnThrVal.....ThrValAlaProAspAlaA 923
1574 TGTTCCTTTATCACCTGGACACGACGCCCTCCCAAAATCCTCTCTGGGAT 1623
923 laAlaValAlaIleThrThrIleAspLeuSerGlyLeu.....Ser 936
1624 CCAGTGGACAGACAGGAGTGGCTTTGTAGTGGACAAAACACCAAAAT 1673
937 AlaThrGlyGlyThrLeuSerGlyThrIleThrHisAsnAlaAlaGlnTh 953
1674 GGCTACCTCCAAATCCAGCATGCTAAGGTTGGCACTTGGAATAACA 1723
953 rAlaLeuThrThrIleLysGly.....960
1724 GTCTGCAAGCAAGCTTCACAAACCTTGACC.....1752
961SerAlaGlyAsnAspThrIleThrIleGlyLysValAsnAspGly 975
1753 CTGACTGTACGTCCTCGCTGCCAAT.....1779
976 LeuThrValThrGlyGlyAlaGlyAsnAspValPheAsnValThrAlaAl 992
1780GCTACCTCGCTCCA...ATTACAGTCACTTCCAAA 1813
992 aLysIleValThrAlaAspThrProGluHisAlaThrIleThrAspPheS 1009
1814 CGAACAGGACACACCAAAATTTCCCGCCCTCTGTTAGTTTATGCAAAAT 1863
1009 eAlaGlyAspSerIleLysPheAlaAlaSerValThrAlaTyrGlyAsn 1025
1864 ATTGCGCAAGAGCGCTCCCAATTCAGGCGCAGTGTACAGCCCTGAT 1913
1026 ValGlyThrValAlaGlyAspThrLeuLysAlaIleLysAlaAlaI 1042
1914 T.....GAAT 1918
1042 eAlaLeuThrAspLysAlaProGlyIleThrSerAlaAspLysGluThrT 1059
1919 CAGTGAATGAAAAACAGTTACTTGGAA.....CTACTGGATAAT 1959
1059 hrValTyrGlyPheThrTyrAsnGlyAspAsnTyrLeuPheTyrAsnAsn 1075
1960 GGAGCAGGTGCTGATCTACTAAG...GATGCGGTGCTCTACTCAAGGTA 2006
1076 AlaAsnGlySerAspSerThrThrValAspValLeuValLysLeuTh 1092
2007 TTTCACAACTTATGACACCAATGGTAGATACAGTGTAAAAGTCGGGCTC 2056


```

1007 AAGCAGGCCAGCTTTCTCCTCTCGACAGACAGTTGAGCTGGGGTCTCGGGTT 1056
      |||:|||||:|||||:
394 .....ThrAlaGlnLeuSerSerSer 401
1057 GGGATGGTGACATTTGACAGCTGCTGCCATGTACAAAGTGAACATCATACA 1106
      :|||
402 LysThrLeuThr.....SerThrGI 408
1107 GATAAACAGTGGCAGTGACAGGACACACTCGCCAAAGATTACCTGCAG 1156
      | :|||||:|||||:
408 nThrThrGlnGlySerPro...SerThrValSerGlnThrThrProGlyV 424
1157 CAGCTTCAGGAGGAGCTCCATCTGCAGCGGGCTTCGATCGGATTTACT 1206
      :|||||:|||||:
424 alSerSerAlaSerThrGlyMetThrThrSerGlnAlaSerLeuArgSer 440
1207 GTGATTAGGAAGAAATATCAACTGATGATCTGAATTCGTGCTGCTGAC 1256
      :|||:|||||:
441 ThrGlnAsnSerValThrSerThrProGlySerPheThrValThrSerTh 457
1257 GGATGGGAAGACAACTATAAGTGGTCTTTAAAGAGGTCAAAACAA 1306
      | :||||
457 rProThrIleThrSerThr.....G 464
1307 GTGGTCCCATCATCCACAGCTCGCTTTGGGGCCCTCTGCAGTCAAGAA 1356
      :|||||:|||||:
464 InGlyAlaMetAlaSerSerSerAsnSerProThrSerThrGlnAla 480
1357 CTAGAGAGCTGCTCCAAAATCAGCAGGAGTTTACAGACATATGCTTCAGA 1406
      :|||||:
481 IleSerSerSerThrValThrThrValAlaSerSerThrIleAlaProTh 497
1407 TCAAGTTTCAGAACAAAGCCCTCATGATGCTTTGGGGCCCTTTTCATCAG 1456
      :|||||:|||||:
497 rSerSerGlnGlySerSerIleAlaSerSerAlaProSerGlnS 514
1457 GAAATGGAGCTGCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGGATTA 1506
      :|||:|||||:
514 erThrThrValIleThrValSerThrThrAlaThrValSerSerGly... 529
1507 ACCCTCCAGAACCCAGCTGGATGAATGGCACAGTATCGTGACAGCAC 1556
      |||:|||||:
530 .....GlnSerThrSerThrPheGlyThrThrIleGlyGlnGlySerSe 544
1557 CGTGGGAAGAGCACTTTGTTCTTATCATCCTGGACAAACGACCCCTCCC 1606
      : |||:|||||:
544 rPheGlySerSerThrIle.....GlnThrThrGlnGlyThrS 557
1607 AAATCCTCTCTGGATCCCACT..... 1629
      : :|||
557 erSerPheAlaProIleProSerThrAlaGlySerSerSerGlnThrPro 573
1630 .....GGACAGAAGCAAGGTGGCTT 1649
      |||||
574 GlySerMetSerSerThrGlyThrThrValGlyGlnMetSerSerSerPh 590
1650 TGTATGGACAAAACACCAAAATG...GCCTACCTCCAAATCCCAAGCA 1696
      | :|||:
590 eGlnProThrAlaProThrSerLeuGlyThrIleMetThrThrProGlyT 607
1697 TTGCTAAGTTGGCACTTGGAAATACAGTCTGCAAGCAACTCACAACCC 1746
      :|||:|||||:
607 hrSerSerIleProSerIleSerThrSerValAsnSerGlySerSerThr 623
1747 TTGACCTGTAGCTGCAGTCCCTGCTGCCAATGCTACC..... 1785
      :|||:|||||:
624 IleGlySerThrValThrGlnAlaProSerSerSerSerSerMetGlyPr 640
1786 .....CTGCCCTCAA 1795
      |||||
640 oSerProSerGlnSerThrAlaGlySerThrMetThrSerAlaProProv 657

```

```

1796 TTACAGTGAAGTTCCTCAAAACGACAAAG...CACACCCAGCAAAATCCCCCAGC 1842
      :|||:|||||:|||||:
657 alThrSerSerSerAlaAsnThrGlySerThrSerSerGlyThrThr 673
1843 CCTCTGTAGTATTCGCAATATTCGCAAGGAGAGCTCCCAATTCCTCAG 1892
      |||:|||||:
674 valSerValGlnThrThrGlnValSerSerThrThrSerProValAlaSe 690
1893 GGCCAGT.....GTCACAGCCCTGATTGAATCATGTAATGGAAAAACAG 1936
      :|||:|||||:
690 rSerSerSerGlnMetThrSer...ThrGlnGlnProSerGlySerSers 706
1937 TTACCTTGGAACTACTGGTAATGCGACAGGCTGCTGATGCTACTAAGCAT 1986
      :|||:|||||:
706 erSerIleGlySerThrValAsnGlnGlySerSerValThrThrGln 722
1987 GACGGTGTCTACTCAAGGTATTTCAACAATTCATGACACAGATGGTAGATA 2036
      :|||:|||||:
723 ProProAlaSerSerArgSerThrAlaSerGlnGlySerSerSerAlaGI 739
2037 CAGTGTAAAGTGGGGCTCTGGAGGAGGTAAACGCA..... 2073
      :|||:
739 nProIleAlaSerSerSerThrMetGlySerThrAlaGlySerSerSerP 756
2074 .....GCCAGACGAGAGTGTATCCCAAGCAGAGTGGAGCA 2109
      :|||:|||||:
756 roGlnProThrAlaSerSerThrSerValProSerSerThrGlyAla 771
seq_name: sp_vertibrate:Q9W6V5
seq_documentation_block:
ID Q9W6V5 PRELIMINARY; PRT; 1406 AA.
AC Q9W6V5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SUPPORTING-CELL ANTIGEN PRECURSOR (EC 3.1.3.48).
OS Gallus gallus (Chicken).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
FN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=INTESTINE;
RC Kugler R.P., Goodyear R.J., Legan P.K., Warchol M., Raphael Y.,
RA Cotanche D.A., Richardson G.P.;
RT "The supporting-cell antigen: a receptor-like protein tyrosine
RT phosphatase expressed in the sensory epithelia of the inner ear.";
RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ238216; CAB41885.1; -.
DR HSSP; P18052; 1YFO.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR00387; TYR_phosphatase.
DR InterPro; IPR000242; Tyr_prot_phptase.
DR Pfam; PF00041; fn3; 9.
DR PRINTS; PR00700; PRTPHPHTASE.
DR SMART; SM00060; FN3; 7.
DR SMART; SM00194; PTPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase; Signal.
FT SIGNAL 1 24 POTENTIAL.
SQ SEQUENCE 1406 AA; 154196 MW; 416387A29B3EC7C2 CRC64;

```

```

alignment_scores:
Quality: 139.00 Length: 556
Ratio: 0.599 Gaps: 21
Percent Similarity: 41.727 Percent Identity: 19.604

```

alignment_block:

US-09-049-696-20 x Q9W6V5 ..
Align seg 1/1 to: Q9W6V5 from: 1 to: 1406

1366 CTGTCCAAATGACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCA 1415
||||| : : : : :
34 LeuLysAsnValThrGluGluMetGlyThrSerSerAsnAspGluLeuSe 50
1416 GAACAATGCCCTCATGTGCTTTTGGGCCCTTTCATCAGGAATGGAG 1465
: ||||| : : : : :
50 rValAsnAla.....ThrSerGlyAsnArga 59

1466 CTGTCTCTCAGCGCTCCATCCAGCTGAGAGTAAGGATTAAACCTCCAG 1515
: : : : :
59 rgLeuSerGlu...AspValSerLeuProGlyArgAlaMetSerAspGln 74

1516 AACAGCCAGTGGATGAATGGCAGACATGATCGTGGACAGCACC.....GT 1559
||||| : : : : :
75 AsnSerValAlaGlnProArgAlaValLeuAspLeuLysThrGluTyrVa 91

1560 GGGAAAGGACACTTGTCTTCTATACCTGGACACGACGCTCCCCAAA 1609
||||| : : : : :
91 lGlyValThrSerValAsnLeu...ThrTrpThrValAsn..... 103

1610 TCCTTCTCTGGGATCCCATGGACAG..... 1635
||| : : : : :
104AspThrAlaSerAspSerTyrThrTyrArgileGluVal 116

1636 AGCAAGGTGGCTTGTACTGGACAAAACCAAAATGGCTACTCTCA 1685
||||| : : : : :
117 ArgAsnGlySerSerileAsnAsnLysThrSerAspIleThrAspAlaG 133

1686 AATCCAGACATGTGTAAGTTGGCACTTGGAAATACATAGCTGCAAGCA. 1734
: ||||| : : : : :
133 uIleThrGlyLeuIleProGlyThrLeuTyrThrPheThrValPheAlav 150

1734 1734

150 alAlaAlaAspGlyGlnThrAlaGlyGluGlyAlaSerileSerLeuty 166
||| : : : : :
1735AG 1736

167 ThrLysProSerProValLeuAspLeuLysAlaGluTyrValGlyValTh 183

1737 CTCACAAACCTTGACCTGACTGTCACGTCCCGTGGTCCCAATGTCACC 1786
||| : : : : :
183 rSerValAsnLeuThrTrpThrValAsnAspThrAlaSerAlaSerTyrT 200

1787 TGCTCCCAATTACAGTCACTTCCAAACCAACAGACACACCAATTC 1836
||| : : : : :
200 hrTyrArgileGluValThrSerAspSerSerileAspSerLeu..... 214

1837 CCCAGCCCTCGTAGTTTATGCAAAATATTCGCCAAGGAGCCTCCCAAT 1886
: : : : :
215ThrSerSerVa 218

1887 TCTCAGGCGCAGTGTACAGCCCTGATTGAATCAGTGAATGGAANAACAG 1936
: : : : :
218 iThrMetAlaGluIleThrGlyLeuIlePro.....GlyThrLeut 232

1937 TTACCTTGGAACTACGTGGATAATGGAGCAGGTGCTGATGCTACTAAGGAT 1986
: : : : :
232 yrSerPheLysValPheAlaValAlaAlaAsp...AsnArgThrGluAla 247

1987 GACGGTGTCTACTCAAGGATTTTACA..... 2013
||||| : : : : :
248 AspGlyAlaSerileSerLeuTyrThrLysProSerProValLeuAspLe 264

2014ACTTATGACACGA 2026
||||| : : : : :
264 uLysAlaGluTyrValGlyValThrSerValAsnLeuThrTrpThrVa 281

2027 ATGGT.....AGATACAGTGTAAAGTGGCGGCTCTC 2058
||||| : : : : :
281 snGlyThrAlaLeuThrAlaTyrThrTyrArgileGluValArgAsnAla 297

2059 GGAGGAGTTAAACGACGACGAGAGTATACCCGACGAGAGTGAGC 2108
: : : : :
298 ThrSerileArgAsnGluThrSerAsnileAsnLysileGluIleThrGI 314

2109 ACTGTACATACCTGGCTGGATTGAGATGATGAATACAAATGAATCCA. 2157
: ||||| : : : : :
314 yLeu...IleProGlyThrSerTyrAsnPheLysValPheAlaThrProv 330

2157 2157

330 alAsnAsnThrThrGluGluGluGlyLeuSerLeuAsnLeuTyrThrLys 346

2158 CCAAGACCTGAAATTAAT..... 2175
||| : : : : :
347 ProSerProValLeuArgValThrGluTyrValGlyValThrSerVa 363

2176AAGGATGATTTCAACACACAAAGTGTGTTTCA 2209
||||| : : : : :
363 lAsnLeuThrTrpMetValAspAspThrAlaSerAspSerTyrThrTyrA 380

2210 GCAGAACATCTCGGAGGCTCATTTCTG.....GCTCTGATGTC 2250
: : : : :
380 rgileGluValArgAsnGlySerSerileAsnAsnLysThrSerAspIle 396

2251 CCAAAATGCTCCCATACCTGATCTCTTCCCA..... 2280
: ||||| : : : : :
397 ThrAspAlaGluIleThrGlyLeuIleProGlyThrLeuTyrThrPheTh 413

2280 2280

413 rValPheAlaValAlaAlaAspGlyGlnThrAlaGlyGluGlyAlaSerI 430

2281CCTGGCAAAATCACCGACCTCAAGCGCGAAAT 2313
||| : : : : :
430 leSerLeuTyrThrLysProSerProValLeuAspLeuLysAlaGluTyr 446

2314 CACGGGGCAGTCTCATTAATCTGACTGGACAGCTCTCTGGGATGATTA 2363
||| : : : : :
447 ValGlyValThrSerValAsnLeuThrTrpThrVal..... 458

2364 TGACCATGGAACAGCTCACAGTATATCATTCGAATAGTACAAGTATTC 2413
: : : : :
459 .AsnAspThrAlaSerAlaSerTyrThrTyrArgileGluValThr.... 473

2414 TTGATCTCAGACAGCAAGTTCAATGAATCTCTCAAGTGAATACT..... 2457
||| : : : : :
474SerAspSerSerileAspSerLeuThrSerAsnValThrMet 487

2459ACTGCTCTCTCCCAAGGAAGCAACTCTGAGGAGTCTT 2498
: : : : :
488 AlaGluIleThrGlyLeuIleProGlyThrLeuTyrAsnPheThrValPh 504

2499 TTGTTTAAACACAGAAAACATTACTTTTGAATAATGGCACAGATCTTTTCA 2548
: : : : :
504 eAlaValAlaAlaAspAsnArgThrGluAlaAspGlyAlaPheThrSerL 521

2549 TTGCTATTTCAGGCTGTTGATAGGTGATCTGCAATCAGAATATCCAAC 2598
: : : : :
521 eutyThrLysProThrProValThrAspLeuLysAlaGlu...HisGly 536

2599 ATGACACGAGTATCTTG 2616
: : : : :
537 ValThrSerValSerLeu 542

OM of: US-09-049-696-20 to: SwissProt_39:* out_format : pfs

Date: Mar 30, 2002 2:57 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+2p.model -DEV=xlp  
-O/cgn2_1/USPRO.spool/US09049696/runat_28032002_145238_2085/app_query.fasta_1.12579  
-DB=SwissProt_39 -QEMT=fastan -SUFFIX=rs -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62  
-TRANS=human40.cdi -DIST=45 -LOCALIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -MINLEN=0 -MAXLEN=200000000  
-USER=US09049696 @CGN1_1_165 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPRY  
-WAIT -THREADS=1
```

Search information block:

Query: US-09-049-696-20

Query length: 2983

Database: SwissProt_39:*

Database sequences: 100059

Database length: 36664827

Search time (sec): 306.030000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
SwissProt_39:Y103_SYNY3	+	171.50	196.46	0.0007	420 ! Q55874 synechocystis sp. (st
SwissProt_39:PMPB_CHAEL	+	153.50	161.59	0.0150	1672 ! Q9piy2 chlamydia muridarum.
SwissProt_39:YPR4_CAMEL	+	136.50	150.28	0.1557	686 ! Q20059 caenorhabditis elegans
SwissProt_39:YPR4_CAMEL	+	136.00	139.92	0.2179	1849 ! P45386 haemophilus influenza
SwissProt_39:ITH2_HESAU	+	135.50	145.93	0.1971	946 ! P97279 mesocricetus auratus
SwissProt_39:PTPD_HUMAN	+	135.50	139.00	0.2372	1912 ! P23468 homo sapiens (human)
SwissProt_39:PTPD_HUMAN	+	135.50	138.59	0.3049	1569 ! P52143 escherichia coli o111
SwissProt_39:EAE_ECOLI	+	133.00	143.10	0.2870	935 ! Q31000 escherichia coli o111
SwissProt_39:PGBM_HUMAN	+	132.00	126.68	0.5017	4393 ! P98160 homo sapiens (human)
SwissProt_39:PMPE_CHLMU	+	131.50	140.90	0.3643	976 ! Q9pi147 chlamydia muridarum.
SwissProt_39:FLIC_ECOLI	+	131.00	146.96	0.3291	497 ! P04949 escherichia coli fla
SwissProt_39:FLIC_SALMU	+	130.50	146.23	0.3563	504 ! P06177 salmonella muenchen.
SwissProt_39:RPEL_DROME	+	129.50	132.05	0.5867	1887 ! P04052 drosophila melanogaster
SwissProt_39:FATH_HUMAN	+	129.00	123.29	0.7412	4590 ! Q14517 homo sapiens (human)
SwissProt_39:NEO1_MOUSE	+	129.00	133.76	0.5950	1493 ! P97798 mus musculus (mouse)
SwissProt_39:FLIC_ECOLI	+	128.50	131.58	0.6697	1754 ! O8418 chlamydia trachomatis
SwissProt_39:FLIC_SALPA	+	128.00	143.49	0.5174	493 ! P06178 salmonella paratyphi
SwissProt_39:VP03_HSVSA	+	128.00	134.36	0.6602	1246 ! Q01000 herpesvirus saimiri
SwissProt_39:FLIC_SALRU	+	127.50	142.92	0.5578	492 ! P06175 salmonella rubislaw.
SwissProt_39:YMG6_YEAST	+	127.50	134.65	0.6957	1140 ! Q04893 saccharomyces cerevis
SwissProt_39:PMPC_CHLMU	+	127.50	132.21	0.7425	1460 ! Q9piy1 chlamydia muridarum.
SwissProt_39:YFBK_ECOLI	+	126.50	140.20	0.6762	575 ! P76481 escherichia coli. hyp
SwissProt_39:FLGK_BORBU	+	126.50	139.35	0.6918	627 ! P70859 borrelia burgdorferi
SwissProt_39:ICG2_HAELN	+	126.50	129.52	0.8995	1702 ! P45384 haemophilus influenza
SwissProt_39:FLIC_HAELN	+	126.00	141.21	0.6990	489 ! P06179 salmonella typhimurium
SwissProt_39:FLJB_SALAE	+	125.50	140.40	0.7584	500 ! P52615 salmonella abortus-ec
SwissProt_39:GUNZ_CLOSAR	+	124.00	131.94	1.114	986 ! P23659 clostridium stercoar
SwissProt_39:MGPC_MYCPN	+	122.50	128.09	1.51	1218 ! Q50341 mycoplasma pneumonia
SwissProt_39:MRP_STPSU	+	122.50	127.78	1.52	1256 ! P32653 streptococcus suis. m
SwissProt_39:CAL7_HUMAN	+	121.50	118.21	2.22	2944 ! Q02388 homo sapiens (human)
SwissProt_39:DYHC_NEUCR	+	121.50	114.33	2.46	4367 ! P45443 neurospora crassa. c
SwissProt_39:FIB1_ADEGI	+	121.00	131.63	1.64	710 ! Q54761 avian adenovirus gall
SwissProt_39:FLGE_HELPJ	+	121.00	131.52	1.65	718 ! Q92k0 helicobacter pylori
SwissProt_39:IGAL_HAELN	+	121.00	124.00	2.02	1541 ! P42782 haemophilus influenza
SwissProt_39:MAC1_YEAST	+	120.50	136.28	1.54	417 ! P35192 saccharomyces cerevis
SwissProt_39:FUGE_CAUCR	+	120.50	132.85	1.69	591 ! P35806 caulobacter crescentu
SwissProt_39:EAE_ECO57	+	120.50	128.34	1.91	934 ! P43261 escherichia coli o157
SwissProt_39:RAD3_SCHPO	+	120.50	119.10	2.44	2386 ! Q02099 schizosaccharomyces
SwissProt_39:FLJB_SALTY	+	120.00	133.80	1.75	505 ! P52616 salmonella typhimurium
SwissProt_39:HXA2_HAELN	+	120.00	127.81	2.05	928 ! P45354 haemophilus influenza

SwissProt_39:ITH2_PIG + 120.00 127.74 2.06 935 ! O02668 sus scrofa (pig). in
SwissProt_39:WAPA_BACSU + 120.00 118.73 2.62 2334 ! Q07833 bacillus subtilis
SwissProt_39:SPG7_DICDI + 119.50 132.70 1.91 532 ! P22698 dictyostellium disc
SwissProt_39:EAE_ECO27 + 119.50 127.10 2.22 939 ! P19809 escherichia coli o1
seq_name: SwissProt_39:Y103_SYNY3

seq_documentation_block:

ID Y103_SYNY3 STANDARD; PRT; 420 AA.

AC Q55874;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE HYPOTHEICAL 45.8 KDA PROTEIN SLLO103.

GN SLLO103.

OS Synechocystis sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

OX NCBI_TaxID=1148;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96127529; PubMed=8590279;

RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,

RA Suglura M., Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium

Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb

region from map positions 64% to 92% of the genome.";

RL DNA Res. 2:153-166(1995).

CC -1% SIMILARITY: TO E.COLI YFBK.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch.

DR EMBL; D64004; BAA10635.1; -

DR InterPro; IPR002035; VMFA.

DR Pfam; PF00092; vwa; 1.

DR SMART; SM00327; VWA; 1.

DR PROSITE; PS0234; VMFA; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 420 AA; 45849 MW; E711B51478E74F3 CRC64;

alignment_scores:

Quality: 171.50 Length: 455

Ratio: 0.776

Percent Similarity: 48.571 Percent Identity: 23.956

alignment_block:

US-09_049-696-20 x Y103_SYNY3

Align Seg 1/1 to: Y103_SYNY3 from: 1 to: 420

775 ACAGACAAACACCAACAAAGCTCCAAACAGCAAAATCAAAATG 824

10 SerAspGlnAsnLeuAspAlaGlyAlaProThrSerGlnArgGln..... 24

825 CAATCTCCGAACACATGGGAAGTGCCTGATCTGAGGACTTTAAGA 874

25LeuArgGileAla.....ValAlaAlaLysAlaAspHisAspA 38

675 AAACCACTCTGATGACACACAGCCACCAATCCACCTTCTCATTGCTG 924

38 rgArgLeuProLeuAsn..... 43

525 CAGATTGGACAAAGAAATGTGTGTTTAGCTCTTGACAAATCTGGAAGCAT 974

44LeuCysLeuValLeuAspHisSerGlySerMe 54

975 GCGGACTGGTAACCGCCTCAATCGACTGAATCAAGCAGGCCACCTTTTCC 1024

```
1 |||... |||... : : : : : : : : : : : : : : |||
54 t...AspGlyGlnProLeuGluThrValLysSerAlaAla...LeuGlyL 69
1025 TGCTGCAGACAGTTGAGCTGGGGTCCCTGGGTGGGATGGGTGACATTTCAC 1074
|||... : : : : : : : : : : : : : : |||
69 euIleAspArgLeuGluGluAspArgLeuSerValIleAlaPheAsp 85
1075 AGTGCTGCCCATGTACAAAGTGAAGTCACTACAGATAAAGAGTGGCACTGA 1124
|||... : : : : : : : : : : : : : : |||
86 HisArgAlaLysIleValIleGluAsnGlnValArgAsnGlyAlaAl 102
1125 CAGGCACACACTCGCCAAAGATTACCTGCAGCAGCTTCAGAGGGACGT 1174
|||... : : : : : : : : : : : : : : |||
102 alleAlaLysAlaIleGluArgLeu...LysAlaGluGlyGlyThra 117
1175 CCATCTGCAGGGCTTCGATCGGCATTTACTGTGATTAGGAAGAATAT 1224
|||... : : : : : : : : : : : : : : |||
117 laIleAspGluGlyLeuLysLeuGlyIleGlnGluAlaLysGlyLys 133
1225 CCAACTGATGATCTGAATTTGCTGCTGACCGATGGGAAGACACAC 1274
|||... : : : : : : : : : : : : : : |||
134 GluAspArgValSerHisIlePheLeuLeuThrAspGlyGluAsnGlu.. 149
1275 TATAAGTGGTGCTTTACAGAGGTCAAAACAAAGTGGTCCCATC... 1317
|||... : : : : : : : : : : : : : : |||
150 ...HisGlyAspAsnAspArgCysLeuLysLeuGlyThrValAlaSerA 165
1318 .....ATCCACACAGCTCGTTTGGGGCCCTCTGCAGCTCAA 1353
|||... : : : : : : : : : : : : : : |||
165 sPtyrLysLeuThrValHisThrLeuGlyPheGlyAspHisTrpAsnGln 181
1354 GAA...CTAGAGGAGCTCTCAAAATGACAGGAGGTTTACAGACATATGC 1400
|||... : : : : : : : : : : : : : : |||
182 AspValLeuGluAlaIleAlaLysAlaSerAlaGlnGlySerLeuSerTyrI 198
1401 TTCAGAT.....CAAG 1411
|||... : : : : : : : : : : : : : : |||
198 egluAsnProSerGluAlaLeuHisThrPheArgGlnLeuPheGlnArgM 215
1412 TTCAGACAATGCGCTCATGATGCTTTTGGGGCCCTTTCATCAGGAAT 1461
|||... : : : : : : : : : : : : : : |||
215 eSerAsnValGlyLeuThrAsnAlaHisLysLeuLeuGluLeuAlaPro 231
1462 GGAGCT.....GTCTCTCAGCGCTCCATCCAGCT 1490
|||... : : : : : : : : : : : : : : |||
232 GlnAlaHisLeuAlaIleValLysProValAlaGln.....ValSerPr 246
1491 TCAGACTAAGGATTAAACCTCCAGACAGCCAGTGGATGAATGCCACAG 1540
|||... : : : : : : : : : : : : : : |||
246 oGluThrMetAspLeuThrValGlnAsnGlnGlyAlaIleGluGluVala 263
1541 TGATCTGGACACACCGCTGGGAAGAGACACTTTCTTTTATCACCTGG 1590
|||... : : : : : : : : : : : : : : |||
263 rgluGlyAspLeuMetThrAspGlnGluArgValLeuLeuAsn... 278
1591 ACAACGACGCTCCCAAAATCTCTCTGGGATCCAGTGGACACAGCA 1640
278 ..... 278
1641 AGGTGGCTTTGTAGTGACAAAACACAAATGGCTACCTCCAA... 1686
279 .....LeuTyrLeuAspGlnL 284
1687 ..ATCCAGGCGATTGCTAAGTTGGC.....ACTTGGAAATACAGCTGTG 1728
|||... : : : : : : : : : : : : : : |||
284 euLeuProGlyGlnHisValIleGlyGlnValcInIleArgTyrAspasp 300
1729 CAAGCAAGCTCACAAACC.....TTGACCCCT 1754
|||... : : : : : : : : : : : : : : |||
301 ProAlaSerGlyGlnThrAsnLeuLeuSerAspProLeuProLeuThrI 317
1755 GACTGTTCAGCTCCCGTGGCTGCAATCTACCTGCTCCCAATTCAGTGA 1804
|||... : : : : : : : : : : : : : : |||
```

alignment_scores:
Quality: 153.50
Ratio: 0.468
Length: 802
Gaps: 34

```
317 eGlnValGlnThrGlnTyrGlnProSerThrAspValGlnValGlnGluS 334
1805 CTTCCAAAACGAACAGGACAGCAAAATCCCCAGCCCTCTGTA... 1851
|||... : : : : : : : : : : : : : : |||
334 erIleLeuThr.....LeuAlaLysTyrArgGlnThrGlnIleAla 347
1852 .....GTTTATCCAAATATTCCCAAGGAGCTCCCAATTTCTCAG 1892
|||... : : : : : : : : : : : : : : |||
348 GluThrLysLeuLysAlaGlyAspArgGlnGlyAlaAlaThrMetLeuG 364
1893 G...GCCAGTGTACAGCCCTG...ATTGAATCAGTGAATGGAAAAACAG 1936
|||... : : : : : : : : : : : : : : |||
364 nThrAlaAlaLysThrAlaLeuGlnMetGlyAspLysAsnGlyAlaThrI 381
1937 TTACCTTGGAACTACTGGAATATGA.....GCAGGTGCTGAT 1974
|||... : : : : : : : : : : : : : : |||
381 le.....LeuGlnThrAsnAlaThrArgLeuGlnSerGlyGluAsp 394
1975 GCTACTAAGGATGAC 1989
|||... : : : : : : : : : : : : : : |||
395 LeuSerGluGlyAsp 399
seq_name: SwissProt_39:PMPB_CHLMU
seq_documentation_block:
ID PMPB_CHLMU STANDARD; PRT; 1672 AA.
AC Q9PJY2;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE OUTER MEMBRANE PROTEIN PMPB PRECURSOR (POLYMORPHIC MEMBRANE
PROTEIN B).
GN PMPB OR TC0694.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MoPn / Nigg.
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1405(2000).
CC 1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
(POTENTIAL).
CC 1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE002338; AAF39510.1; -.
CC TIGR: TC0694; -.
DR InterPro: IPR003357; OMP.
DR Pfam: PF02385; OMP; 1.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 14 POTENTIAL.
FT CHAIN 15 1672 PROBABLE OUTER MEMBRANE PROTEIN PMPB.
SQ SEQUENCE 1672 AA; 176295 MW; 4ABF190DA4DF8BD6 CRC64;
```


Percent Similarity: 40.898 Percent Identity: 18.080

alignment_block:

US-09-049-696-20 x PMPB_CHLMU ..

Align seg 1/1 to: PMPB_CHLMU from: 1 to: 1672

```
547 TTCTACTTATCCAAATGGAGAATACAGCAGTAGAAGTGTTCACAGGTAT 596
      :::::  :::::  |||  :::::  :::::  :::::  :::::  :::::
122 PheTyrAsnGluHisSerGlyIleLeuSerPheMetAlaArgSerGlyVa 138
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
597 TACTGCT.....ACAAATAGTAAAGAAAGTGTACAGGAGGCA 634
      |||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
138 iGluGlySerLeuThrLeuSerAsnIleLysMetThrGlyAspGlyGlyA 155
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
635 GCTGTATACCAAAA...AGATGCACATTCATAAATAAGTAACAGACTCTAT 681
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
155 laIleTyrSerGlnGlyGluLeuPheThrAspLeuThrGlyLeuThr 171
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
682 GAAAGAGGATGTGAGTTGTCTCCCAATCCCGCCAGACGAGAGGCTTC 731
      :::::  :::::  |||  :::::  :::::  :::::  :::::  :::::
172 IleGlnGly.....AsnLeuSerGlnLeuSerGlyGlyG1 183
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
732 TATAATGTTTGCAACATGTTGATCTATAGTTGAATTCGTACAGAAC 781
      :::::  :::::  :::::  |||  :::::  :::::  :::::  :::::
183 yIlePheGlyGly.....SerThrIleSerPhe..... 192
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
782 AAACCAACAACAAAGAAGCTCCAAACAGCAAAATCAAAATCAATCTC 831
      :::::  :::::  :::::  :::::  |||||  :::::  :::::  :::::
193 .....SerGlyIleAsnGlnAlaThrPheSer 201
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
832 CGAAGCACATGGGAAGTATGCTGATCTTACGAGCTTTAAGAAACACC 881
      :::::  :::::  |||||  :::::  :::::  :::::  :::::  :::::
202 SerAsnThrAlaGluValProGlu.....GluThrTh 213
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
882 TCCT.....ATGCAACACACACCCACCAATCCCA 910
      |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
213 rProAsnProAsnProGlyThrGlnThrThrSerGlnProSerProT 230
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
911 CTTCTCATTTGTCAGATTGGACAAGAATGTGTGTTTATGCTTGTGAC 960
      ||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
230 hr.....SerLysValGlnSerLeuPheThrTyr 239
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
961 AAATCTGGAGCATGGCACTGCTACCGCTCAATCGACTGATCAAGC 1010
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
240 SerSerSerThrGlnAlaAsnGlyAsn..... 248
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1011 AGCCAGCTTTTCTCTGTCGACAGACAGTTGAGCTGGGGTCTGGGTGGGA 1060
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
248 ..... 248
1061 TGCTGACATTTGACAGTGTGTCCTGATGACAAAGTGAATCATACAGATA 1110
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
248 ..... 248
1111 AACAGTGGCAGTCACAGGACACACATCGCCAAAGATTACCTCCAGCAGC 1160
      |||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
249 .....GlyAlaAspSerGlnThrProSerHisLys.....ProG1 260
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1161 TTCAGGAGGAGCATCCATCTGCACGCGGGTTCGATCGGCATTTACTGTGA 1210
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
260 ySerGlyGlyAlaIleTyrAlaThrGly.....AspLeuThrIle. 273
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1211 TTAGGAAGAAATATCAACTGATGATCTGAATTTGCTGCTGACGGAT 1260
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
274 .....SerAspSerGlnGluIleValPhe..... 281
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1261 GGGGAAGACACACACTATAAGTGGTCTTTAACGAGGTCAACAAAGTGG 1310
      |||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
282 .....SerValAsnLysAlaSerLysAspG1 290
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1311 TGCCATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAACTAG 1360
      |||  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```

```
290 yGlyAlaIle..... 293
1361 AGGAGCTGTCCAAATGACAGAGGTTTACAGACATATGCTTCAGATCAA 1410
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
294 .....PheAlaGluLysAsnValSerPheGluAsnIleThrThrLeuLys 308
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1411 GTTCAGAACAAATGGCCTCATTTGATGCTTTTGGGGCCCTTTCATCAGGAAA 1460
      |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
309 valGlnAsnAsn.....GlyAlaGluGluLysGlyG1 319
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1461 TGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGATTAAGGATTAAACCC 1510
      |||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
319 yGlyIleTyrAlaSerGlyAspLeuSerIleGlnSerSerLysGlnSerL 336
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1511 TCAGAACACGCGCAGTGGATGAATGCGACAGTGAATGCTGGACAGCAGCGTG 1560
      :::::  ||  |||||  :::::  :::::  :::::  :::::  :::::  :::::
336 euPheAsnSer..... 339
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1561 GGAAGGACACATTTGTTCTTATCACCTGGCAACGCGCTCCCAAAAT 1610
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
339 ..... 339
1611 CTTCTCTGGATCCCGCTGGACAGCAAGCTGGCTTTGTAGTGGACA 1660
      :::::  |||  |||  |||  :::::  :::::  :::::  :::::
340 .....AsnThrSerLysGlnGlyGlyAlaLeuTyrIleGluG 353
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1661 AAACACCAAAATGGCTACCTCCAAATCCAGGCAATGCTAAAGTTGGC 1710
      |||  :::::  :::::  |||  :::::  :::::  :::::  :::::
353 lyAsnValAspPheLysAspLeuGluIleArgIle..... 365
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1711 ACTTGAATATACAGTCTGCNAGCAAGCTCACAAACCTTGACCTGACTGT 1760
      |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
366 .....LysTyrAsnLysSerGlyThrPheGluThrLysLysValThrLe 380
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1761 CAGCTCCGCTGCCAATGCTACCTCCCTCCCAATTACAGTCACTTCCA 1810
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
380 uSerLeuProGluAlaGlnThrAsnLysSerSerValThrAlaAlaSerG 397
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1811 AAACGACAGGACACAGCAAAATCCCGAGCCCTCTGGTA..... 1851
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
397 lnSerGlyProAsnThrThrProThrProThrProValThrAlaLys 413
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1852 .....GTTTATGCAATATTCGCAAGGAGCCTCCCAATTCCTCAG 1892
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
414 GlyGlyGlyLeuTyrThrGluLysAsnLeuSerIleSer..... 426
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1893 GGCACGTGTCACAGCCCTGATTGAATCAGTGAATGGAATAACAGTTACCT 1942
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
427 .....AsnIleThrGlyIleIleGluIleThrAsnAsnLysAlaThr. 440
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1943 TGGAACTACTGGATAATGGAGCAGGTGCT..... 1971
      |||  |||  |||  |||  |||  |||  |||  |||
441 .....AspValGlyGlyAlaTyrValLysGlyThrLeuThr 453
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1972 .....GATGCTACTAA 1982
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
454 CysLysAspSerHisArgLeuGlnPheGlnLysAsnSerSerGluLysLy 470
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1983 GGATACGGTGTCTACTCAAGGTATTTCACAACTTATGACACGAATGGTA 2032
      |||  |||  |||  |||  |||  |||  |||  |||
470 sGlyGlyGlyLeuTyr.....ThrGluAspThr..... 479
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2033 GATACAGTGTAAAGCTCGCGCTCTGGAGGAGTAAACGACGACACAGCG 2082
      :::::  :::::  |||  |||  |||  |||  |||  |||
480 .....IleThrLeuSerAsnLeuThrGlyLysThrLeuPheGlnGlu 493
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2083 AGAGTGATACCCAGCAGAGTGGAGCTGTACATACCTGCCTGGATTGA 2132
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
494 AsnThrAlaLysGluGlyGlyGlyLeuTyrIleGlnGly..... 507
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2133 GAATGATGAAATACAATGGAATCCCAACAGCCTGAAATTAATAAGGATG 2182
      |||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
508 .....Aspa 509
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```


DR MEROPS; S06.001; -
DR InterPro; IPR000710; Iga_S6.
DR Pfam; PF02395; IGA1; 1
DR PRINTS; PRO0921; IGASERTFAS.
KW Hydrolyase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1022 1849 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 299 299 PROBABLE.
FT SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;

alignment_scores:
Quality: 136.00 Length: 956
Ratio: 0.346 Gaps: 53
Percent Similarity: 41.109 Percent Identity: 21.339
alignment_block:
US-09-049-696-20 x IGA4_HAEIN ..
Align seg 1/1 to: IGA4_HAEIN from: 1 to: 1849

349 TCTACTCTCCAGCT.....AATGATGAACCTACACTGAGCAG..... 387
279 SerAspProLysGlyLeuSerGlnAspProLeuThrAsnTyrAlaVa 295
388 ATGGGCAACTGTGA.....GAGAAGGCTG 412
295 lLeuGlyAspSerGlySerProLeuPheValTyrAspArgGluLysGlyL 312
413 AAAGATCCACCTC...ACTCCTGATTTCATTCGAGGAAAGATAGCT 459
312 ysrTrpLeuPheLeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLys 328
460 GAATATGACACACAGGTAGGCGCTTCCTCATGAGTGGGCTCATCAG 509
329 SerTrp...GlnGluTrpAsnIleTyrLysHisGluPheAla..... 341
510 ATGGGGAGTATTGACGAGTACAATAATGATGAGAAATTCATTATCCA 559
342GluLysIleTyr..... 345
560 ATGGAGATACACAGCAGTAAGATGTTCCAGCAGGT...ATTACTGGTACA 606
346GlnGlnTyrSerAlaGlySerLeuThrGlySer 356
607 AATGTAGTAAAGAAGTGTGACGGAGGAGCTGTTACACCAAAAGATGCAC 656
357 Asn.....ThrGlnTyrThrTrpGlnAlaTh 365
657 A.....TTCAATAAGTAACAGGACTCTATGCAAAAGATGTGAGTTG 700
365 rGlySerThrSerThrIleThrGly..... 373
701 TTCTCCAATCCCGCAGGAGGAGGCTTCTATATGTTTGACAAACAT 750
374GlyGlyGluProLeuSerValAspLeuThrAspGly 385
751 GTTGATTCATATGTAATCTGTACAGCAACAAACCAACAAAGAC 800
386 LysAsp.....LysProAsnHisGlyLysSerI 395
801 TCCAAACACAAATCAAAATGCAATCTCCGA..... 834
395 eThrLeuLysGlySerGlyThrLeuThrLeuAsnAsnHisIleAspGlnG 412
834 834
412 lYAlaGlyGlyLeuPhePheGlyAspTyrGluValLysGlyThrSer 428
835AGCATGGGAAGTATCCGTGATTCTGAGGACCTTTAAGAAAC 878
429 AspSerThrThrTrpLysGlyAlaGlyValSerValAlaAspGlyLysTh 445

879 CACTCCTATGACACACAGCCACCAAAATCCACCTCTCTCA...TTGCTGC 925
445 r.....ValThrTrpLysValHisAsnProLysTyrAspArgLeuAlaL 460
926 AGATTGGACAAAGAATTGTGTGTGTAGTCCCTTGAC.....AAA 963
460 ysileGlyLysGly.....ThrLeuValValGluGlyLysGlyLysAsn 474
964 TCTGAAGCATGGCAGCTGGTAAACCGCTCAATCGACTGAATCAACAGCAGG 1013
475 GluGlyLeuLeuLysValGlyAsp.....G 483
1014 CCAGCTTTTCCTGCTGCAG.....ACAGTTGAGCTGG 1045
483 yThrValIleLeuLysGlnLysAlaAspAlaAsnAsnLysValGlnAlaP 500
1046 GTCTCTGGTGTGGATGTGACATTTTGACAGTGTGCTGCCCATGTACAAAGT 1095
500 heSerGlnValGlyIleValSerGlyArgSerThr..... 511
1096 GAATCATACAGATAACAGTGGCAGTGCAGGACACACTGCCCAAAG 1145
512LeuValLeuAsn.....AspAspLysGlnValAspProAsnSe 524
1146 ATTACCTGCAGCAGCTTTCAGGAGGAGCTCCATCTGCAGCGGCTTCGAT 1195
524 rIleTyrPheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnSerL 541
1196 CGCATTATTACTGTGATTAGGAAGAATATCCAACCTGATGATCTGAAAT 1245
541 euThrPheAspHisIleArgAsn.....IleAspAspGlyAlaArgVal 555
1246 GTG.....CTGCTGACGATGGGGAAGACACACTATAAGTGGG..... 1284
556 ValAsnHisAsnMetThrAsnThrSerAsnIleThrIleThrGlyLys 572
1284 1284
572 rLeuIleThrAsnProAsnThrIleThrSerTyrAsnIleGluAlaGlnA 589
1285TGC 1287
589 sPAspAspHisProLeuArgIleArgSerIleProTyrArgGlnLeuTyr 605
1288 TTTAAACAGCTCAACAAAGTGTGCTCATCCACACAGCTGCTTTGGG 1337
605 PheAsnGlnAspAsnArgSer.....TyrTyrThrLeuLysLysG 619
1338 GCCTCTGTCAGCTCAAGACTAGAGGAGCTGTCCAAATGACA..... 1380
619 YAlaSerThrArgSerGluLeuProGlnAsnSerGlyGluSerAsnGluA 636
1381GGAGTTTACAGACATATGCTTCAGATCAAGTTCAG 1416
636 snTrpLeuTyrMetGlyArgThrSerAspGluAlaLysArgAsnValMet 652
1417 AAC.....AATGGCTCATTCATGCTTTTGGGGC 1445
653 AsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrPheGly.. 668
1446 CCTTTCATCAGGAATGGAGTGTCTCTCAGCGCTCCATCCAGCTTGAGA 1495
669GluGluGluThrLysAlaThrGlnAsnGlyLysLeuAsnValThrP 684
1495 GTAAGGATTAACCTCCAGAACAGCCAGTGGATGAATGGCAGCAGTGATC 1545
684 heAsnGlyLysSerAspGlnAsnArgPheLeuLeuThrGlyGlyThrAsn 700
1546 GTG.....GACAGCAGCTGGGAAAGACACTTTGTTCTTATCACTGT 1589
701 LeuAsnGlyAspLeuAsnValGluLysGlyThrLeuPheLeuSerGlyAr 717

378 CACTGACGAGATG.....GCCAATCTGGGAGACAAGGGTGAAGAGTACC 421
 : :::||||| ::: ||||:: ::: ||||:: ||||
182 styrGlnGluMetLysTrpArgLysLeuGlySerTyrrGluHisLysIleH 199

422 ACCTCACTCCTCATTCATTGCAGCAAAAAGAATAGCTGAATATGGACCA 471
 ||||| ||| :::||| ||| ||| :::
199 isLeuGlnProGlyArgLeuAla.....LysHisLeuGlValAsnVal 213

472CAAAGTAGGGCATTGTGTCATGAGTGCGGCTCATCT 506
 ||||| ||||| ||||| |||||
214 TrpIleValGluLeuGlnGlyMetArgPheLeuHis..... 225
507 ACGATGGGAGATTTGACGAGTACAANTAAATGATCAGAAATTCATCTAT 556
 ||| ||| ::: :::
226ValProAspThrPhe..... 230

557 CCAAATGGAAGATAACAAGCATAAGA...TGTTCCAGCAGGTATTACTGGT 603
 :::|||||:::|||||::: ||| |||
231 .GluGlyHisPheGlnGlyValProValIleSerLysGlyGlnLysLys 246
604 ACAATGTGTAGTAAGAAG.....TGTCAGGAGCAGCCTGTTACAC 644
 :::|||||::: ||| :::||||| ::: ||||
247 SerHisValSerPheLysProThrValAlaGlnArgLysCys.... 261
645 CAAAAAGATGCACATTCAAATAAGATACAGACTCTATGAAAAAGGATGTG 694
 :::|||||::: ||| ||| ::: ||||
262 ProAsnCysThrTyrrThrAlaValAspGly..... 271
695 AGTTTGTCTCCAATCCCGCAGCAGGAGAAGGCTTCATATGTTTGC 744
271ValProAspThr..... 271
745 CAACATGTTGATTCTATAGTTGAATTCTGTACAGACAAAAACCACAACAA 794
 :::|||||::: ||| ::: ||||
272GluLeuValValMetTyrrAspValAsnAr 281
795 AGAAGTCCAACACAGCAAAATCAAAATGCAATCTCCGACGACATGGG 844
281 .ggLu.....G 283
845 AAGTGAATCCGTGATTCTGAGGACTTTAAGAAAAACCCTCCTATGACAACA 894
 || ::: ||| |||||
283 LylsValGlyGluLeuGluValPheAsnGlyTyrrPheValHisPhePhe 299
895 CAGCCCAAAATCCACCTTCTCATTTCTGCAGATTGGACAAGATTCT 944
 ||| ||| ||||| ::: ||||
300 AlaProGluAsn.....LeuaspProIleproLysasnIl 311
945 GTGTTTAGTCCTTGACAAATCTGGAAGCATGGGACTGTTAACCGCTCA 994
 : :::|||||::: ||||| ||| ::: ||||
311 eLeuPheValIleaspValSerGlySerMet...TrpGlyIleLysMetL 327

995 ATCACTAGTATCAAGCAGGCCAGCTTTCTGTG.....CTGCAGACAGTT 1038
 Gln:::||||| ::: ||||
327 ysGlnThrValGluAlaMetLysThrIleLeuAspLeuArgThrGlu 343
1039 GAGCTGGGGTCTCGGTTGGGATGGTGACATTTGCACAGTCTGCCCATGT 1088
 ||| ||| ::: |||| ||| ::: ||||
344 AspGlnPheSer.....ValValaspPheAsnHisAsnValargTh 357

1089 A...CAAAGTCAACTCATACAGATAAACACTGGCAGTGACAGGACACAC 1135
 : :::|||||::: ||||
357 rTrpArgasnAspleuValSerAlaThrLysThrGlnIleThrAsp.... 372
1136 TCGCCCCAAAGATA.....CCTGCAGCACCTTCAGGAGGAGCGTCCAT 1179
 ||||| ::: ||||
373 ...AlaLysArgTyrrIleGluLysIleGlnProSerGlyGlyThrAsnIle 388

1180 TGCACCGGGCTTCGATCGGCATTTACTGTGATTATAGGAGAAA..... 1221
 :::||||| ||| ::: ||||
389 AsnGluAlaLeuLeuArgAlaIlePheIleLeuAsnGluAlaSerAsnLe 405

1222TATCCAATGATGATCTGAAATTTGCTGCTGACGGATG 1261

```

alignment_scores:      Length: 1008
      Quality: 135.50      Gaps: 54
      Ratio: 0.320      Percent Identity: 20.139
Percent Similarity: 42.063

alignment_block:
US-09-049-696-20 x ITH2_MESAU ..

Align seg 1/1 to: ITH2_MESAU from: 1 to: 946

82 CTGAGTAATTCATCATTCAG.....CTGAACAACAATGCCTATGAGG 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
85 MetAlathrThrIleLeGlnSerLysLeuValAsnAsnSerProGlnSe 101
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
126 CATTGCTGTTGCATCGAACCCCAATGTGCCAGGAAGATGAA..... 165
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
101 rGlnAsnValAlpheaspValGlnIleProLysGlyAlaPheIleSera 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
166 .....ACATCATTCACAACAATAAAGAC 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
118 snPheThrMetThrValAsnGlyIleThrPheThrSerThrIleArgGlu 134
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
190 ATGCTGACCCAGGCATCTCTGAT...CTGTTGAAGCTACAGGAAGACG 236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
135 LysThrValGlyArgAlaLeuTyrSerGlnAlaargAlaLysGlyLys.. 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
237 ATTTTATTCAAAATGTTGCCATTTTGATTTCCTGAAACATGGAAGACAA 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
151 .....T 151

287 AGGCTGACTATCTGAGACCAAACTTGAGACCTACAAAAATGCTCATGTT 336
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
151 hrAlaGlyTrpValArgSerArgThrLeuAspMetGluAsnPheAsnThr 167
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
337 CTGCTTGCTGAGTCTACTCTCCACAGT.....AATGATGAACCCCTA 377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
168 GlnVal.....AsnIleProGlnValAlaLysValGlnPheGluLeuHi 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

405 uGlyMetLeuAsnProAspSerValSerLeuLeuValleValSerAspG 422
1262 GGAAGACAAACATATAAGTGGGTCTTTAAGGAGGTCAACAAAGTGGT 1311
422 ly...AspProThrVal...GlyLeuLeuLys... 430
1312 GCATCATCCACACAGTCCTTTGGGGCCCTCTGCAGCTCAAGAACTAGA 1361
430 430
1362 GGAGCTGTCCAAAATGACAGGAGTTTACAGACATATGCTTCAGATCAAG 1411
431LeuSerLys...IleGlnLysAsnValLysGlnAsnI 442
1412 TTCAGAACAAATGGCCTCATGTATGCTTTTGGG... 1443
442 leGlnAspAsnIleSerLeuPheSerLeuGlyIleGlyPheAspValAsp 458
1444GCCCTTTCATCAGAAATGGAGCTGTCTCTCAGCG 1478
459 TyrAspPheLeuLysArgLeuSerAsnGluAsnArgGlyIleAlaGlnAr 475
1479 CTCCATCCAGCTTGAGAGTAAGGATTAACCTCCAGAACACAGCAGTGA 1528
475 gile...TyrGlyAsnArgAspThrSerSerGlnLeuLysLysPheT 490
1529 TGAATGGCACAGTGTGCGGACAGCCTGGGAAAGGACACTTTGTTT 1578
490 yrAsnGln...ValSerThrProLeuLeuArgAsnValGlnPhe 503
1579 CTATCATCCTGCAGACGAGCTCCCAAAATCCTCTCTCGGATCCCAAG 1628
504AsnTyrProGlnAlaSerValThrAsp... 512
1629 TGGCAGAGAAGAGTGGCTTTGTAGTGCAACAAACACAAATGGCCT 1678
513ValThrGlnAsnSerPheHisAsnT 521
1679 ACCTCAAAATCCAGGCAATGTGTAAGTTGGCACTTGGAAATACAGTCTG 1728
521 yrPheGlyGlySerGluIleValValAlaGlyLysTyrAspProSerLys 537
1729 CAAGCAAGCTCAAACTTGACCTGACCTGACCTGACCTCCCGTCCGTC 1778
538 LeuAlaGluValGlnSerIle...IleThrAlaThrSerThrAsnThrG 553
1779 TGCTACCTCGCTCCCAATTACA...GTGACTT 1807
553 uLeuValLeuGluThrLeuSerGlnMetAspAspLeuGluAspPheLeu 570
1808 CCAAAACGAACAGGACACCAATCCCGAGCCCTCTGCTAGTTTAT 1857
570 erLysAspLysHisAlaAspProAsnPheThrLysLysLeuTrpAlaTyr 586
1858 GCAATATTCGCCAA...GGAGCCTCCCAATCTCTCAG 1892
587 LeuThrIleAsnGlnLeuLeuAlaGluArgSerLeuAlaPro... 600
1893 GGCCAGTGTCCAGCCCTGATTGAATCAGTGAATGGAAAACAGTT... 1938
601ThrAlaAlaIleLysArgLysIleThrLysThrIleLeuG 614
1939ACCTTGGAACTACTG 1953
614 InMetSerLeuAspHisIleValThrProLeuThrAlaMetValIle 630
1954 GATAATGGACAGGCTGCTGATCTACTAGGATGACGGTCTCTACTCAAG 2003
631 GluAsnGluAlaGlyAspGluArgMetLeuAlaAspSerPro... 644
2004 GTATTTCACAACTTATGACACCAAGTGGTAGATACACTGTAAAAGTCCGG 2053

644 644
2054 CTCTGGGAGGAGTTAAACGACGCCAGCGAGAGTGATATCCCGACAG 2100
645ProGlnAspHis 648
2101ATGGAGCACTGTAC... 2115
649 SerCysCysSerGlyAlaLeuTyrTyrGlyThrLysValAlaSerAla 665
2116 ATACTGTGGTGG... 2127
665 rIleProSerTrpAlaSerProSerProThrProValMetAlaMetLeu 682
2128 ..ATTGAGAATGATGAATACAAATGAATCCACCAAGACCTGAATTAAT 2175
682 laValGlyAlaAsnArgLeuGluSerThrProProHisValIleArg 698
2176 AAGGATGATGTTCAACAC...AAGCA 2198
699 ValGluAsnAspProHisPheIleIleTyrLeuProLysSerGlnLysAs 715
2199 AGTGTGTTTTCAGCAGAACATCTCGGGAGGCTCATTGTGGTCTCTGATG 2248
715 nIleCysPheAsnIleAspSerGlu... 723
2249 TCCCAAAATGCTCCCATACCTGATCTCTTCCACCTGGCCAAATCACCAG 2298
724ProGlyLysIleLeuSer 729
2299 CTGAGGCGGAAATTCACGGGGGCGCAGTCTCAATTAATCTGACTGGACAG 2348
730 LeuValSerAspProGluSerGlyIleLeuValAsn...GlyGlnLeuI 745
2349 TCCTGGGATGATTATGACCATGGAACAGCTCACAAAGTATATC... 2391
745 eGlyAlaLysLysAlaGluAsnGlyLysLeuArgThrTyrPheGlyLysL 762
2392ATTGGAATAAGTACAAAT 2409
762 euGlyPheTyrPheGlnLysGluAspMetLysIleGluIleSerThrGlu 778
2410 ATTCTGTATCTCAGACACAAGTTCAATGAATCTCTTCAAGTGAATACTAC 2459
779 AsnIleThrLeu...IleAsnGlySer...SerThrTh 789
2460 TGCTCTCATCCCAAGGAAGCC...AAGCTGAGGAAGTCTTTT 2500
789 rSerLeuPheTrpSerAspThrAlaHisLeuGlyAsnGlnArgValLeuI 806
2501 TGTTT...AAACCAGAAAATTTACTTTTGAATAATGGCAGATCTT 2544
806 leSerValLysLysGlyLysSerValThrLeuThrLeuAsnLysGluMet 822
2545 TTCATTCTATTTCAGGCTGTTGATAAGTCGATCGAATCAGAAATATC 2594
823 PhePheSerVal...LeuLeuHisValTrpLysLysHisProVal... 837
2595 CAACATTCGACGAGTATCTTTGTTTATCTCCACAGACTCCGCCAGAGA 2644
838 AsnValAspPheLeuGlyIleTyrLeuProPro...ThrAsnLysPheS 853
2645 CACCTAGTCTGATGAACAGCTGCTGCTCTGT...CCTAATATT 2685
853 erProSerAlaHisGlyLeuLeuGlyGlnPheMetAsnLysProAsnIle 869
2696 CATATCAACAGCACCATTCTCTGGC 2709
870 HisIlePheAsnGluArgProGly 877
seq_name: SwissProt_39:PTPD_HUMAN
seq_documentation_block:

ID PTPD_HUMAN STANDARD; PRT; 1912 AA.
 AC P23468;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROTEIN-TYROSINE PHOSPHATASE DELTA PRECURSOR (EC 3.1.3.48) (R-PTP-
 DELTA).
 GN PTPRD.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND MUTAGENESIS OF ARG-1178.
 RX MEDLINE=95204468; PubMed=7896816;
 RA Pulido R., Krueger N.X., Serra-Pages C., Saito H., Streuli M.;
 RT "Molecular characterization of the human transmembrane protein-
 tyrosine phosphatase delta. Evidence for tissue-specific expression of
 alternative human transmembrane protein-tyrosine phosphatase delta
 isoforms.";
 RT J. Biol. Chem. 270:6722-6728(1995).
 RL [2]
 RN SEQUENCE OF 390-1912 FROM N.A.
 RP TISSUE=Placenta;
 RX MEDLINE=91006018; PubMed=2170109;
 RA Krueger N.X., Streuli M., Saito H.;
 RT "Structural diversity and evolution of human receptor-like protein
 tyrosine phosphatases.";
 RT tyrosine phosphatases.";
 RL EMBL J. 9:3241-3252(1990).
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
 PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS ARE FOUND IN DIFFERENT
 TISSUES DUE TO ALTERNATIVE SPLICING.
 CC -1- PPM: A CLEAVAGE OCCURS THAT SEPARATES THE EXTRACELLULAR DOMAIN
 FROM THE TRANSMEMBRANE SEGMENT.
 CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 8 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 DR EMBL: L38929; AAC41749.1; -;
 DR EMBL: X54133; CAA38068.1; -;
 DR PIR: S12052; S12052.
 DR HSP: F18052; LYFO.
 DR MIM: 601598; -;
 DR InterPro: IPR001777; FN_III.
 DR InterPro: IPR003006; Ig_MHC.
 DR InterPro: IPR003598; Ig_C2.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR InterPro: IPR000242; Tyr_prot_phptase.
 DR Pfam: PF00041; fn3; 8.
 DR Pfam: PF00047; Ig; 3.
 DR Pfam: PF00102; Y_phosphatase; 2.
 DR PRINTS: PR00014; FNTYPEIII.
 DR PRINTS: PR00700; PRTPHPHTASE.
 DR SMART: SM00060; FN3; 8.
 DR SMART: SM00408; IGc2; 3.
 DR SMART: SM00194; PTPC; 2.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 2.
 DR PROSITE: PS00566; TYR_PHOSPHATASE_2; 2.
 DR PROSITE: PS00555; TYR_PHOSPHATASE_PTP; 2.
 KW hydrolase; Receptor; Glycoprotein; Signal; Transmembrane; Duplication;
 FT Immunoglobulin domain; Alternative splicing.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 1912 PROTEIN-TYROSINE PHOSPHATASE DELTA.

FT DOMAIN 21 1265 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1266 1290 POTENTIAL.
 FT DOMAIN 1291 1912 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 23 115 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 118 225 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 232 318 IG-LIKE C2-TYPE DOMAIN 3.
 FT DOMAIN 320 414 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 417 513 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 516 606 FIBRONECTIN TYPE-III 3.
 FT DOMAIN 609 708 FIBRONECTIN TYPE-III 4.
 FT DOMAIN 711 822 FIBRONECTIN TYPE-III 5.
 FT DOMAIN 825 916 FIBRONECTIN TYPE-III 6.
 FT DOMAIN 918 1017 FIBRONECTIN TYPE-III 7.
 FT DOMAIN 1020 1137 FIBRONECTIN TYPE-III 8.
 FT DOMAIN 1375 1618 PROTEIN-TYROSINE PHOSPHATASE 1.
 FT DOMAIN 1619 1912 PROTEIN-TYROSINE PHOSPHATASE 2.
 FT ACT_SITE 1553 BY SIMILARITY.
 FT ACT_SITE 1844 BY SIMILARITY.
 FT SITE 1175 CLEAVAGE (POTENTIAL).
 FT CARBOHYD 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 724 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 832 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 181 MISSING (IN KIDNEY ISOFORM).
 FT VARSPLIC 226 MISSING (IN KIDNEY ISOFORM).
 FT VARSPLIC 775 MISSING (IN KIDNEY ISOFORM).
 FT VARSPLIC 609 MISSING (IN FETAL BRAIN ISOFORM).
 FT MUTAGEN 1178 R->A: 2.5-FOLD REDUCTION IN CLEAVAGE.
 SQ SEQUENCE 1912 AA; 214759 MW; 3A8BCBCD32182E26 CRC64;

alignment_scores: Length: 757
 Quality: 135.50
 Ratio: 0.420
 Percent Similarity: 42.668
 alignment_block:
 US-09-049-696-20 x PTPD_HUMAN ..
 Align seg 1/1 to: PTPD_HUMAN from: 1 to: 1912
 670 ACAGGACTCTATGAAGAAGATGTGATTGTTCTCAATCCGCCGACGAC 719
 |||||:
 34 ThrGlyValSerGlyGlyValAlaSerPheileCysGlnAlaThrGlyAs 50
 720 GGAGAAGCTTCATATATGTTTGCACAAACATGTTGATTCTATAGTTGAAT 769
 :
 50 pProArgProLysileValTrp..... 57
 770 TCTGTACAGCAAAACACACAAAGAGCTCCAAACAGCAAAATCAA 819
 |||||:
 58AsnLysLysGlyLysLysValSerAsnGln 67
 820 AAATGCAATCTCCGAAGCAGATGGAAAGTATCGGT.....GATTC 860
 :
 68 Arg.....PheGluValIleGluPheAspaspGlySe 78
 861 TGAGGACTTTAAGAAACCACTCTCTATCACACACACGCCCAATCCCA 910
 | :
 78 rGlySerValLeuArgIleGlnProLeuArgThrProArgAspGluAlaI 95
 911 CCTTCTCAATGCTGCAGATGGACAAAGAATGTGTGTAGTCTCTTGAC 960
 :
 95 leTyrGlu.....CysValAlaSerAsn 102
 961 AAATCTGGAAGCATGGGACGTGGTAACCGCTCAATCACTGAATCAAGC 1010
 :
 103 AsnValGlyGluIleSerValSerThrArgLeuThrValLeuArgGluAs 119
 1011 AGGCCAGCTTTTCTCTGTCGACAGACAGTGTGAGCTGGGTGGGTGGGA 1060
 :
 119 pGlnIleProArgGlyPheProThrIleAspMetGlyProGlnLeuLysV 136

1061 TGGTGACATTTGACAGTGTGCCATGTACAA..... 1092
136 aValGluArgThrArgThrAlaThrMetLeuCysAlaAlaSerGlyAsn 152
1093 1118
153 ProAspProGluIleThrTrpPheLysAspPheLeuProValAspThrSe 169
1119 CAGTGACGGGACACACTCCCAAGATTACTGCAGCAGCTTCAGGAG 1168
169 rAsnAsnGlyArgIle...LysGlnLeuArgSerGluSerIleGlyC 185
1169 GGAGGTCCTATCGCAGCGGCTCGATCGGCATTACTGTGATTAGGAAG 1218
185 lYthrProIle.....ArgGlyAlaLeuGlnIleGluGlnSer 197
1219 AAATATCCAACTGATGATCTGAAATTTGCTGCAGCGATGGGAAGA 1268
198 GluGluSerAspGlnGlyLysThrGluCysValAlaThrAsnSerAlaGl 214
1269 CAACACTATAAGTGGTGTCTTAAC..... 1293
214 yThrArgTyrSerAlaProAlaAsnLeuTyrValArgGluLeuArgGluV 231
1294GAGTTC 1299
231 alArgArgValProProArgPheSerIleProProThrAsnHisGluIle 247
1300 AAACAAGTGGTCCATC...ATCCACACAGTCGCTTGGGGCCCTCT... 1344
248 MetProGlyGlySerValAsnIleThrCysValAlaValGlySerProMe 264
1345GCAGCTCAAGAACTAGAGGAGCTGTCCA 1372
264 tProTyrValLysTrpMetLeuGlyAlaGluAspLeuThrProGluAspA 281
1373 AAATGACGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGACAAAT 1422
281 spMetProIleGlyArgAsnValLeuGluLeuAsnAspValArgGlnSer 297
1423 GGC.....CTCATTTGATGCTTTTGGGGCCCTTTCAFCAGG 1457
298 AlaAsnTyrThrCysValAlaMetSerThrLeuGlyValIleGlu..... 312
1458 AAATGAGCTGTCTTCACGGCTCCATCCAGCTTGAGTAGAGGATTA 1507
313AlaIleAlaGlnIleThrValLysAlaLeuProLys..... 324
1508 CCCTCCAGAACAGCCAGTGGATGAATGGCACA...GTGATCGTGGACGC 1554
325ProProGlyThrProValThrGluSer 334
1555 ACCGTGGGAAGGACACTTTGTTCTTATCAGCTGGACACGACGCTCC 1604
335 ThrAla.....ThrSerIleThrLeuThrTrpAspSerGlyAsnPr 348
1605 CCAATCTCTTCTCGGATCCAGTGGACAGAGCAAGTGGCTTTGTAG 1654
348 oGluProValSerTyr.....TyrIleI 356
1655 TGGAC.....AAAAACACCAAAATGGCTTACTCCAAATCCAGGC 1695
356 lGlnHisLysProLysAsnSerGluGluLeuTyrLysGluIleAspGly 372
1696 ATTGCTAAGTGGCACTTTGGAAATACAGTCTGCAAGCAAGCTCACAAA 1745
373 ValAla..... 374
1746 CTTGACCCCTGACTGCACGTCGCGTCCAAATGCTACCCCTGCCPCCAA 1795
375ThrThrArgTyrSerValAlaGlyLeuSerProT 386
1796 TTACAGTGACTTCCAAACGACACAGGACACCAAGCAAAATCCCGCCCT 1845

386 yz.....SerAspTyrGluPheArg 392
1846 CTGGTAGTTTATGCAATATTCGCCAAGGAGCTCCCAATTTCTCAGGCG 1895
393 ValValAlaValAsnAsnIleGlyArgGlyProPro..... 404
1896 CAGTGTACAGCCCTGATTGAATCAGTGAATGGAAAAACAGTTACCTGG 1945
405SerGluProValLeuThrGln...ThrSerGluGlnAlaProSer 418
1946 AACTACTGGATAATGGAGCAGGTGTGCTACTAAGGATGACGCGTGC 1995
418 SerAlaProArgAspValGlnAlaArgMetLeuSerSerThrThrIleLe 435
1996 TACTCAAGGTATTTTCAACAATTATGACCAATGGTAGATACAGTGATA 2045
435 uValGln..... 437
2046 AGTGGGGCTCTGGGAGGAGTTAACCCAGCAGACGAGAGTGATACCCC 2095
438TrpLysGluProGluGluProAsnGlyGln..... 447
2096 AGCAGAGTGGAGCAGCTGTACATACCTGGCTGGATTGAGATGATGAATA 2145
448IleGlnGlyTyrArgValTyrTyrT 456
2145 CAATGGAATCCACCAAGACCTGAAATTAAT.....AAGGATGATGTCA 2189
456 hrMetAspProThrGlnHisValAsnAsnTrpMetLysHisAsnValAl 472
2190 ACACAAAGCAAGTG..... 2202
472 aAspSerGlnIleThrThrIleGlyAsnLeuValProGlnLysThrTyrS 489
2203TGTTCAGCAGACATCTCGGAGGCTCATTTGGGCTTCT 2244
489 erValLysValLeuAlaPheThrSerIleGlyAspGlyProLeuSerSer 505
2245 GATGTCCCAAAATGCTCCATACCTGCTCTTCCACCTGGCCAAATCAC 2294
506 AspIle.....GlnValIleThrGlnThrGlyValProGlyGlnProLe 520
2295 CGACCTGAAGGGGAAATTCACGGGGCAGTCTCATTAATCTGACTGGA 2344
520 uAsnPheLysAlaGluProGluSerGluThrSerIleLeuLeuSerTrpT 537
2345 CAGCTCTGGGATGATTATGACCATGGACAGCTCACAGTATATCAT 2394
537 hrProProArgSerAsp.....ThrIleAlaAsnTyr..... 547
2395 CGAATAAGTACAAGTATTTCTGATCTCAGAGACAAGTTCAATGAATCTCT 2444
547 547
2445 TCAAGTGAATACTACTGCTCTCATCCAAAGNAGCAACTCTGAGAAG 2494
548GluLeuValTyrLysAspGlyGluHisGlyGlu. 558
2495 TCTTTTGTGTTAAACCAGAAACATTACTTTGAAAATGSCACAGAT... 2541
559GluGlnArgIleThrIleGluProGlyThrSerTyr 570
2542CTTTTCATTGCTATTTCAGGC 2561
571 ArgLeuGlnGlyLeuLysProAsnSerLeuTyrTyrPheArgLeuAlaI 587
2563 TGTGTATAAGTGTGATCTGAATATCAGAAATATCCAACTTCACAGATAT 2611
587 aArgSerProGlnGlyLeuGlyAlaSerThrAlaGluIleSerAlaArgT 604
2612 CTTTGTGTTTATCTCCACAGACTCCGCCAGAG.....ACACCTAGT 2652
.....

[illegible][illegible]

seq_name: SwissProt_39:EAE_ECOL1

seq documentation block:

seq_documentation_block:	
ID	EAE ECOLL STANDARD: PRT: 935 AA:

ID	LAN_ECC
AC	0371000;

AC 031000;
DT 20-AUG-DT 20-AUG-
DT 20-AUG-20'-AUG-
DT

DT 20-AUG-78
DE 1201PM

DE INTIMIN
GN FAF OR

GN ESCHERICHIA OR ESCHERICHIA

Escheri
bacteri

OC	Bacteri
OC	Bacteri

OC	Escheri
OC	NC-PT
OC	TC

OX	NCBI_Ta
OW	(1)

8N [1]

RP	SEQUENC
----	---------

RX	MEDLINE
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

RA VOSS E.

RT	"Molecu
----	---------

Rf protein

RT Syndrom

RL Ir.fect.

CC - ! - FUN

CC
CC

CC - i - SUE

CC - i - SIM

CC
CC

2000

CC This SW

between
this study

between

• • •

100

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF025311; AAC69247.1; ..
CC InterPro; IPR003344; Big_1.
CC InterPro; IPR003343; Big_2.
CC InterPro; IPR003335; Intimin.
CC InterPro; IPR002482; LysM.
CC Pfam; PF02369; Big_1; 2.
CC Pfam; PF02368; Big_2; 1.
CC Pfam; PF01476; LysM; 1.
CC PRINTS; PR01369; INTIMIN.
CC SMART; SM00257; LysM; 1.
KW Outer membrane; Virulence.
FT REPEAT 65 113 LYSM.
SQ SEQUENCE 935 AA; 101570 MW; 406E79CD07DEB11 CRC64;

alignment_scores:
Quality: 133.00 Length: 572
Ratio: 0.465 Gaps: 30
Percent Similarity: 50.000 Percent Identity: 20.979

alignment_block:
US-09-049-696-20 x EAE_ECOL1 ..

Align seg 1/1 to: EAE_ECOL1 from: 1 to: 935

1096 GAACATCATACAGATAAAGTGC.....AGTGACAGGGA 1130
:::|||||:::|||||:::|||||:::|||||:::|||||
436 AspLeuValGlnArgAsnAsnAlaLeuLeuGluTyrLysGlnAs 452
1131 CACATCGCGCAAAAGATTACCTGCACAGCTTCAGGAGGAGCTCCATCT 1180
|:::|||||:::|||||:::|||||:::|||||
452 pIleLeuSerLeuAsnAlaProHisAspLeuGluHisSerT 469
1181 GCAGCGGCTTCATCGGCATTTACTGTGATTAGGAAGAAATCAACT 1230
:::|||||:::|||||:::|||||:::|||||
469 hrGlnLysIleGln.....LeuIleValLysSerLysTyrGlyLeu 482
1231 GATGGATCTGAATTTGCTGCTGCAGGATGGGGAAGACAACTATAAG 1280
|||:::|||||:::|||||:::|||||
483 Asp.....ArgIleVal.....TyrAspAspSerAlaLeuAr 493
1281 TGGTGCTTTAACAGAGTCAAAAGTGGTGCATCATCCACACAGTCG 1330
:::|||||:::|||||:::|||||:::|||||
493 g.....SerGlnGlyGlyGlnIleGlnHisSer..... 502
1331 CTTTGGGCGCTCTGCAGCTCAAGAACTAGAGGAGCTG.....TCCAA 1374
|||:::|||||:::|||||:::|||||
503GlySerGlnSerAlaGlnAspTyrGlnAlaIleLeuProAlaTyr 517
1375 ATGACAGAGTTTACACACATATGCTTCAGATCAAGTTCAGACAAATGG 1424
:::|||||:::|||||:::|||||
518 ValGlnGlyGlySerAsnIleTyrLys..... 526
1425 CCTCATGATGCTTTTGGGCGCTTTTCATCAGGAATGGAGCTCTCTCTC 1474
:::|||||:::|||||:::|||||
527ValThrAlaArgAlaTyrAspArgAsnGlyAsn.....SerS 539
1475 AGCGCTCCATCCAGCTTGAGAGTAAGGATTACCTCCAGACAGCCAG 1524
:::|||||:::|||||:::|||||
539 erAsnAsnValGlnLeuThrIle.....ThrValLeuSerAsnGlyGln 553
1525 TGGATGATGCCAGATGATCGTCGACAGCACCGTGGGAAG..... 1566
:::|||||:::|||||:::|||||
554 ValValAspGlnValGlyValThrAspPheThrAlaAspLysThrSerAl 570
1567GACACTTTGTTCTTATCACCTGGACAAACGACGCTCCCAAA 1609
|||:::|||||:::|||||:::|||||

570 alyAlaAspGlyThrGluAlaIleThrTyrThrAla..... 582
1610 TCCTTCTCTGGATCCAGTGGACAGAAGCTGGCTTTGTAGTGGAC 1659
:::|||||:::|||||:::|||||
583ThrValLysLysAsnGly.....ValThrGln 591
1660 AAAACACCAAAATGGCTACCTCCAAATCCCAAGGCAATGCTAAGTTGG 1709
|||:::|||||:::|||||:::|||||
592 AlaAsnValProValSerPheAsnIleValSerGlyThrAlaThrLeuG 608
1710 CACTTGAATACAGTCTGCAACAGCTCACAAACCTTGACCCCTGACTG 1759
|||:::|||||:::|||||:::|||||
608 yAlaAsnSerAlaThrThrAspAlaAsnGlyLysAla.....ThrV 622
1760 TCAGTCCCGTGGTCCCAATGCTACCTGCTCCCAATTCACAGTACTCC 1809
|||||:::|||||:::|||||:::|||||
622 alThrLeuLysSer.....ThrProGlnValValSerAla 636
1810 AAAACGAACAGGACACCAATTTCCCGAGCCCTCTGCTAGTTATGTC 1859
|||||:::|||||:::|||||:::|||||
637 LysThrAlaGluMetThrSerAlaLeuAsnAlaSerAlaValIlePheVa 653
1860 AATATTTCGCCAAGGAGCTCCCAATTTCTCAGGCGCAGTGTCCACGCC 1909
:::|||||:::|||||:::|||||:::|||||
653 lGluGlnThrLysAlaSerIleThrGluIleLysAlaAspLysThrThra 670
1910 TGATTGAATCAGTGAAT...GGAAAAACAGTTACCTTGGAACTACTGGAT 1956
:::|||||:::|||||:::|||||:::|||||
670 laValAlaAsnGlyAsnAspAlaValThrTyrThrValLysValMetLys 686
1957 AATGACAGGCTGCTGATGCTACTAAGTATGACGCTCTCTACTCAAGGTA 2006
:::|||||:::|||||:::|||||
687 GluGlyGlnProValHisGly.....HisSerValAl 697
2007 TTTCCAACTTAT.....GACAGCAATGTTAGATACAGTGTAAAGTGC 2050
|||||:::|||||:::|||||:::|||||
697 aPheThrThrAsnPheGlyMetPheAsnGlyLysSerGlnThrGlnAsnA 714
2051 GGGCTCTGGGAGGATTAACCCAGCCAGGAGAGTATACCCAGCAG 2100
:::|||||:::|||||:::|||||
714 laThrThrGlySerAspGlyArgAlaThrIleThrLeuThrSerSerSer 730
2101 AGTGAGCAGCTGTACATACCTGGCTGAGTGAATGATGAATACAAATG 2150
:::|||||:::|||||:::|||||
731 AlaGlyLysAlaThrValSerAlaThrValSerGlyGlyAsn..... 744
2151 GAATCCACCAAGCTGAAAT.....A 2173
745 AspValLysAlaProGluValThrPhePheAspGlyLeuLysIleAspA 761
2174 ATAAGGATGATGTTCAACACAGCAAGTGTGT..... 2205
|||||:::|||||:::|||||
761 snLysValAspIleLeuGlyLysAsnValThrGlyAspLeuProAsnIle 777
2206TTACAGCAACATCCTCGGAGGCTCATTTGT 2237
778 TrpLeuGlnTyrGlyGlnPheLysLeuLysValSerGlyGlyAsnGlyTh 794
2238 GCCTTCT...GATGTCCAAATGCTCCCATACCTGATCTCTCCACCTG 2284
:::|||||:::|||||:::|||||
794 rTyrSerTrpHisSerGluAsnThrAsnIleAlaThrValAspGluSerG 811
2285 GCCAATACCCAGCTGAGGCGGAAATTCACGGGGGAGTCTCATTAAT 2334
|||||:::|||||:::|||||:::|||||
811 lLysValThr...LeuLysGlyLys.....GlyThrAlaValIleAsn 824
2335 CTGACTTGGACAGCTCTCGGGATGATTATGACCATGGA.....AC 2375
:::|||||:::|||||
825 ValThr.....SerGlyAspLysGlnThrValSerTyrThrIlely 838
2376 AGCTCAAGATATATCATTCGAATAGTACAAAGTATTCTTGATCTCAGAG 2425
|||:::|||||:::|||||
838 sAlaProAsnTyrMetIleArgValGly..... 847

[illegible]

RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
 RA Tryggvason K.;
 RA Cloning of human heparan sulfate proteoglycan core protein,
 RT assignment of the gene (HSPG2) to lp36.1-->p35 and identification of
 RT a BamHI restriction fragment length polymorphism.";
 FL Genomics 11:389-396(1991).
 [5]
 RN SEQUENCE OF 1-21 FROM N.A.
 RN MEDLINE 94052171; PubMed 8234307;
 RX Cohen I.R., Graessell S., Murdoch A.D., Iozzo R.V.;
 RT "Structural characterization of the complete human perlecan gene and
 RT its promoter";
 RT Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
 RL CC -1: FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT
 CC MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
 CC CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
 CC PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
 CC SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
 CC SUBSTRATE FOR CELLS.
 CC -2: SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
 CC DIMERS OR STELLATE STRUCTURES.
 CC -3: SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -4: TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
 CC -5: PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
 CC AND O-LINKED OLIGOSACCHARIDES.
 CC -6: SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
 CC SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
 CC -7: SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
 CC -8: SIMILARITY: CONTAINS 22 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -9: SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
 CC -10: SIMILARITY: CONTAINS 1 EGF-LIKE DOMAINS.
 CC -11: SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC -12: THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@lsb-sib.ch).

 DR EMBL; X62515; CAA44373.1; -;
 DR EM3L; M85289; AAS52700.1; -;
 DR EM3L; M64283; AAS52699.1; -;
 DR EMBL; S76436; AAB21121.2; -;
 DR EMBL; L22078; -; NOT_ANNOTATED_CDS.
 DR HSSP; P00740; 11XA.
 DR MIM; 142461; -;
 DR InterPro; IPR0000561; EGF-like.
 DR InterPro; IPR0000742; EGF_2.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR021272; LDL_recept_A.
 DR InterPro; IPR0000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR000082; SEA.
 DR Pfam; PF000008; EGF; 4.
 DR Pfam; PF00047; Ig; 22.
 DR Pfam; PF00052; laminin_B; 3.
 DR Pfam; PF00053; laminin_EGF; 8.
 DR Pfam; PF00054; laminin_G; 3.
 DR Pfam; PF00057; ldl_recept_a; 4.
 DR Pfam; PF01390; SEA; 1.
 DR -PRINTS; PR00010; EGFBL00D.
 DR ProDom; PD003031; Laminin_B; 3.
 DR SMART; SM00180; EGF_Lam; 6.
 DR SMART; SM00001; EGF_like; 8.
 DR SMART; SM00408; IGC2; 22.
 DR SMART; SM00281; LamB; 3.
 DR SMART; SM00282; LamG; 3.
 DR SMART; SM00192; LDLa; 4.

SMART: SM00200; SEA; 1.
 PROSITE; PS00022; EGF_1; 9.
 PROSITE; PS01186; EGF_2; 5.
 PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
 PROSITE; PS01209; LDLRA_1; 4.
 PROSITE; PS00068; LDLRA_2; 4.
 PROSITE; PS50024; SEA; 1.
 Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
 Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
 Extracellular matrix; EGF-like domain.
 SIGNAL 1 21
 CHAIN 22 4393
 DOMAIN 80 194
 DOMAIN 197 236
 DOMAIN 283 321
 DOMAIN 323 361
 DOMAIN 366 405
 DOMAIN 405 506
 DOMAIN 523 532
 DOMAIN 533 732
 DOMAIN 733 765
 DOMAIN 766 815
 DOMAIN 816 873
 DOMAIN 881 925
 DOMAIN 926 935
 DOMAIN 936 1127
 DOMAIN 1128 1160
 DOMAIN 1161 1210
 DOMAIN 1211 1267
 DOMAIN 1277 1326
 DOMAIN 1327 1336
 DOMAIN 1337 1531
 DOMAIN 1532 1564
 DOMAIN 1565 1614
 DOMAIN 1615 1672
 DOMAIN 1679 1773
 DOMAIN 1774 1867
 DOMAIN 1868 1957
 DOMAIN 1958 2053
 DOMAIN 2054 2153
 DOMAIN 2154 2246
 DOMAIN 2247 2342
 DOMAIN 2343 2438
 DOMAIN 2439 2535
 DOMAIN 2536 2631
 DOMAIN 2632 2728
 DOMAIN 2729 2828
 DOMAIN 2829 2926
 DOMAIN 2927 3023
 DOMAIN 3024 3114
 DOMAIN 3115 3213
 DOMAIN 3214 3300
 DOMAIN 3301 3401
 DOMAIN 3402 3490
 DOMAIN 3491 3576
 DOMAIN 3577 3671
 DOMAIN 3672 3847
 DOMAIN 3848 3883
 DOMAIN 3886 3924
 DOMAIN 3966 4104
 DOMAIN 4106 4143
 DOMAIN 4145 4178
 DOMAIN 4243 4391
 SITE 65 67
 SITE 71 73
 SITE 76 78
 SITE 4151 4153
 SITE 4301 4303
 SITE 199 212
 SITE 206 225
 DISULFID 199 212
 DISULFID 206 225

```
1812 .....HisAsnGlyLysLeuProThrArgAlaMetAspPheAsnGlyIle 1826
1966 GGTGCTGATCTACT.....AAGATGACGGTGTCTAC..... 1998
1827 LeuThrIleArgAsnValGlnLeuSerAspAlaGlyThrTyrValCysTh 1843
1999 .....TCAAGGTATTTCACACTTATGACAGCAATGGTAGATACAGTAA 2044
1843 rGlySerAsnMetPheAlaMet.....AspGlnGlyThrAlaThrLeuH 1858
2045 AAGTCGGGCTCTGGGAGGAGTTAACGCAGCCAGCAGAGATGATACCC 2094
1858 isValGlnAla..... 1861
2095 CAGCAGAGTGGAGCAGCTGATACATCTGGCTGGATTGAGAAATGANAAT 2144
1862 .....SerGlyThrLeuSerAlaPro.....ValVa 1870
2145 ACAATGAATCCACCAAGACCTGAAATTAATAAGGATGATGTTCAACACA 2194
1870 iSerIleHisProProGlnLeu.....ThrValGlnProG 1882
2195 AGCAAGTGTGTTTCAGCAGAACATCCTCGGGAGGCTCATTTGTGGCTTCT 2244
1882 lyGlnLeuAlaGluPheArgCysSerAlaThrGlySer..... 1894
2245 GATGTCCCAATGCTCCATACCTGATCTC.....TTCACACC 2282
1895 .....ProThrProThrLeuGluThrGlyGlyProG 1906
2283 TGCCCAATCACCACCTGAGCGGGAATTCACGGGGGAGCTCATTA 2332
1906 yGlnLeuLeu...ProAlaValAlaGlnIleHisGlyGly...IleLeuA 1921
2333 ATCTGACTGGACAGCTCTCGGGATGATATGAC..... 2367
1921 rgLeuProAlaValGluProThrAspGlnAlaGlnTyrLeuCysArgAla 1937
2368 CATGGAACAGCTCACAGTATATCATTCGTAATAGTACAGTATCTTGA 2417
1938 HisSerSerAlaGlyGlnGlnValAlaArg.....AlaValLeuH 1951
2418 TCTCAGACAGAAGTTCAATGAATCTCTCAAGTGAAT..... 2454
1951 svalHisGlyGlyGlyProArgValGlnValSerProGluArgThrG 1968
2455 .....ACTACTGCTCTC..... 2466
1968 lnValHisAlaGlyArgThrValArgLeuTyrCysArgAlaAlaGlyVal 1984
2467 .....ATCCCAAGGA 2477
1985 ProSerAlaThrIleThrTrpArgLysGluGlySerLeuProProG 2001
2478 AGCCAACCTCTGAGAA.....GNCCTTTTGTGTTAAACACAGAAACA 2518
2001 nAlaArgSerGluArgThrAspIleAlaThrLeuLeuIleProAlaIle 2018
2519 TTACTTTTGAATGGCAGACATCTTTTCATTCGTAFTACAGGCTGTTCAT 2568
2018 hrThrAlaAspAlaGlyPheTyrLeuCysValAlaThrSerProAlaGly 2034
2569 AAGTCGATCTGAATCAGAAATATCCAAATTCAGCAGAGATATCTTTGTT 2618
2035 ThrAlaGlnAlaArgIleGlnValValLeuSerAlaSerAspAlaSe 2051
2619 TATTCTCCACAGCTCCGCCAGACACCTAGTCCCTGATCAACAGCTCTG 2668
2051 rGlnProPro...ValLysIleGluSerSerProSerValThrGluG 2067
2669 CTCCTTCTCCTAATATTCATATCAACAGCACCATTCTGGCATTCACATT 2718
```

```
2067 ly.....GlnThrLeuAspLeuAsnCysValValAlaGlySerAlaHis 2081
2719 TTAATAATTATGTGG 2733
2082 AlaGlnValThrTrp 2086
seq_name: SwissProt_39:PMPE_CHLMU
seq_documentation_block:
ID PMPE_CHLMU ; STANDARD; PRT; 976 AA.
AC Q9PL47;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE OUTER MEMBRANE PROTEIN PMPE PRECURSOR (POLYMORPHIC MEMBRANE
DE PROTEIN E);
GN PMPE OR TC0261.
OS Chlamydia muridarum.
OC Racteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RL Nucleic Acids Res. 28:1397-1406(2000).
CC 1. SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC 2. SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC 3. THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL: AE002293; AAF39130.1;
DR TIGR; TC0261;
DR InterPro: IPR003357; OMP.
DR Pfam: PF02385; OMP; 1.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 976 PROBABLE OUTER MEMBRANE PROTEIN PMPE.
SQ SEQUENCE 976 AA; 105964 MW; 0BEA56158E3CEE9 CRC64;
alignment_scores:
Quality: 131.50 Length: 861
Ratio: 0.353 Gaps: 39
Percent similarity: 43.206 Percent identity: 18.118
alignment_block:
US-09-049-696-20 x PMPE_CHLMU
Align seg 1/1 to: PMPE_CHLMU from: 1 to: 976
238 TTTTATTTCAAAAATGTTGCCATTTTGTATCTCTGAAACATGGAAGACAAA 287
95 PhecysPheLysasn.....LeuThrProGluSerGlyGlyVal11 108
288 GGCTGACTATGTGAGA.....CCAAACTTGAGACCTACAAAATGCTG 331
108 eGlySerProThrGlnAsnThrProThrIleLysIleHisAsnThrIleG 125
332 ATGTTCTGTTGCTGAGTCT.....ACT 354
```

333 snHsthrglyGlyAlaIleTyrIleThrGlyThrGluThrSer..... 347

1207 GTGATTAGGAAGAAATATCAACTGATGATGACTGAAATTTGGTGTGTCAC 1256
||||| :|||

348LysIleSerAlaPheHisAlaIlePhe..... 358

1257 GGATGGGAACACACACACTATAAGTGGGTCTTTAACGAGGTCAACAAA 1306
|||||:|||||:|

359AspAsnIleSerAla.....AsnAlaThrAsnAlaA 370

1307 GTGTGTCATCATCCACACA..... 1326
:|||||:

370 spGlySerSerSerAsnThrAsnProHisArgAsnAlaIleThrMet 386

1327GTGGCTTTGGGGCCCTCTGCAGCTCAGAAGCT 1358
:|||||:|

387 AspAsnSerAlaGlyGlyIleGluLeuGlyAlaGlyLysSerGln..... 401

1359 AGAGGAGTGTCCAAATGACAGGAGTTTACAGACATATGCTTCAGATC 1408
:|||||:|

402AsnLeuIlePheTyrAspProIleG 410

1409 AAGTTCAGAACATGGCCTCATTTGATGCTTTTGGGGCCCTTTTCATCAGA 1458
||||| ||| :|||:|

410 InValThrAsnAlaGlyValThrValAspPheAsnLysAspAlaSerGln 426

1459 RATGAGCTGCTCTCAGCGCTCATCCAGCTTCAGAGTTCAGAGTAAGGGATTAA 1508
:|||||:|

427 ThrGlyCysValVal..... 431

1509 COTCCAGAACACCGCAGTGGATGATGGCAGAGTATCGTGGACAGCACCG 1558
:|||||:|

432PheSerGlyAlaThrValLeuSerAlaAspI 442

1559 TGGGAAGGACACTTTGTTCTTATCACCTGGCAACGCGACTCCCCAA 1608
:|||||:|

442 IeSerGlnAlaAsnLeu.....GlnThrLysThrProAlaThr 454

1609 ATTCCTTCTCGGATCCCAAGTGGAGACAGAGTGGCTTTGTAGTGA 1658
:|||||:|

455 LeuThrLeu.....SerHisGlyLeuLeuCysIleG 465

1659 CAAAACACCAAAATGGCC.....TACCTCCAAATCCCAAGCGATG 1699
:|||||:|

465 uAspArgAlaGlnLeuThrValAsnAsnPheThrGlnThrGlyGlyIleV 482

1700 CTAAGGTTGGC.....ACTTGGAAATACAGCTCTGCAA 1731

482 aAlaLeuGlyAsnGlyAlaValLeuSerSertyrGlnHisSerThrThr 498

1732 GCNAGCTCAACAACCTTCACCTGACTGTCAGTCGCGTCCGCTCC..... 1776

499 AspAlaThrGlnThrProProThrThrThrThrThrAspAlaSerValTh 515

1777AATGCTACCTCGCTCCCAATTCACAGTACACTTCCAAA 1813
:|||||:|

515 rLeuAsnHisIleGlyLeuAsnLeuProSerIleLeu..... 527

1813 CGAACAGACACCGACCAATATCCCGCCCTCTGGTAGTTATGCAAAAT 1863
|||||:|

528LysAspGlyAlaGluMetProLeuLeuThrPValGluProIleSer 542

1864 ATTCGCCAAGGAGCTCCCAATTCACAGGCGCAGTGTACAGACGCTGAT 1913
:|||||:|

543 ThrThrGlnGly.....AsnThrThrThrThrThrSerAspThrAlaAlaSe 558

1914 TGAATCAGTCAATPGGAAAAACAGTTACCTTGGAACTGATGGAATCGAG 1963
|||||:|

558 rPheSerLeuAsnGly.....AlaThrLeuSerLeuIleAsp..... 570

1964 CAGGTGCTGATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCACA 2013
:|||||:|

571GluAspGly.....AsnSerProTyrGluAsn 579


```
2014 ACTTATCACAGCAATGGTAGATACAGTGTAAAGCTCGCGGCTCTGGGAGG 2063
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
580 ThrAspLeuSerArgAlaLeuTyrAlaGlnProMetLeuAlaIleSerG1 596
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2064 AGTTAAGCAGCCAGCGGAGTGTATACCCAGCAGAGTGGAGCACTGT 2113
: : : : : : : : : : : : : : : : : : : : : : : : : : :
596 uAlaSerAspAsnGlnLeuGlnSerGluSerMetAspPheSerLysVala 613
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2114 ACATACCT.....GGCTGGATTGAG 2133
: : : : : : : : : : : : : : : : : : : : : : : : : : :
613 snValProHisTyrGlyTrpGlnGlyLeuTrpThrTrpGlyTrpAlaLys 629
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2134 AATGATGAATACATAATGAATCCACCAAGA.....CCTGAATTAATAA 2177
: : : : : : : : : : : : : : : : : : : : : : : : : : :
630 ThrGluAsnProThrThrProAlaThrIleThrAspProLysLys 646
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2178 GGATGATGTTCAACAGCAAGCAAGTGTGTTTCAGCAGCAACATCCTCGGAG 2227
: : : : : : : : : : : : : : : : : : : : : : : : : : :
646 sAlaAsnGlnPheHisArgThrLeuLeuLeuThrTrpLeuProAlaIle 662
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2228 GCTCATTTGTGGTCTGTGTGTCCTCCAAATGCTCCCAT..... 2265
: : : : : : : : : : : : : : : : : : : : : : : : : : :
663 .....TyrIleProSerProLysHisLysSerProLeuIleAlaAsnThr 677
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2266 .....CCTGATCTTCCACCTGCGCAAAATCACCAGCCTGAAGGC 2306
: : : : : : : : : : : : : : : : : : : : : : : : : : :
678 LeuTrpGlyAsnIleLeuPheAlaThrGluAsnLeuLysAsnSerSerG1 694
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2307 GGAA.....ATTCACGGGGGCGAGTCTCA 2329
: : : : : : : : : : : : : : : : : : : : : : : : : : :
694 yGlnGluLeuLeuAspArgProPheTrpGlyIleThrGlyGlyLeuG 711
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2330 TTAATCTGACTGACAGCAGCTCTGGGATGATTATGAC..... 2367
: : : : : : : : : : : : : : : : : : : : : : : : : : :
711 lyMetMetValTyrGlnGluProArgLysAspHisProGlyPheHisMet 727
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2368 .....CATGGAACACTCACAAGTA 2387
: : : : : : : : : : : : : : : : : : : : : : : : : : :
728 HisThrSerGlyTyrSerAlaGlyMetIleThrGlyAsnThrHisThrPh 744
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2388 TATCATTCGAATAGTACAAATGATTCTTGATCTCAGAGCAAGTTC.... 2433
: : : : : : : : : : : : : : : : : : : : : : : : : : :
744 eSerLeuArgPheSerGlnSerTyrThrLysLeuAsnGluArgTyrAlaL 761
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2434 .....AATGAATCTCTCAAGTGAATATCTACTGCT 2463
: : : : : : : : : : : : : : : : : : : : : : : : : : :
761 yAsnTyrValSerSerLysAsnTyrSerCysGlnGlyGlu.MetLeuLe 777
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2464 CTATCCCAAGGAAGCAACTCTGAGGAAG 2494
: : : : : : : : : : : : : : : : : : : : : : : : : : :
777 uSerLeuGlnGlyGlyLeuMetLeuThrLys 787
: : : : : : : : : : : : : : : : : : : : : : : : : : :
seq_name: SwissProt_39:FLIC_ECOLI
```

```
seq_documentation_block:
ID FLIC_ECOLI STANDARD; PRT; 497 AA.
AC P04949;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FLAGELLIN.
GN FLIC OR FLAF OR HAG OR B1923.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=87057066; PubMed=3536885;
RA Kuwajima G., Asaka J.-I., Fujiwara T., Fujiwara T., Node K., Kondo E.;
RT "Nucleotide sequence of the hag gene encoding flagellin of
Escherichia coli.";
RL J. Bacteriol. 168:1479-1483(1986).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=89281489; PubMed=2659972;
RA Hanafusa T., Sakai A., Tominaga A., Enomoto M.;
RT "Isolation and characterization of Escherichia coli hag operator
mutants whose hag48 expression has become repressible by a Salmonella
H1-repressor.";
RL Mol. Gen. Genet. 216:44-50(1989).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sampei G., Seki Y., Sivasubramanian S., Tagami H.,
RA Takada J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 160-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
[5]
RP SEQUENCE OF 1-19 FROM N.A.
RX MEDLINE=83238225; PubMed=6305924;
RA Szekely E., Simon M.;
RT "DNA sequence adjacent to flagellar genes and evolution of flagellar-
phase variation.";
RL J. Bacteriol. 155:74-81(1983).
[6]
RP SEQUENCE OF 1-12.
RC STRAIN-K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313(1997).
CC -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC -!- FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M14358; AAA23950.1; -
CC EMBL; X17440; CAA35488.1; -
CC EMBL; AE000285; AAC74990.1; -
CC EMBL; D90832; BAA15744.1; -
CC EMBL; D90833; BAA15751.1; -
CC EMBL; J01607; AAA92491.1; -
CC PIR; A37249; FLEC.
CC SWISS-2DPAGE; P04949; COLI.
CC EcoGene; EG10321; flic.
CC InterPro; IPR001029; Flagellin_C.
CC InterPro; IPR001492; Flagellin_N.
CC Pfam; PF00700; Flagellin_C; 1.
```

DR Pfam: PF00669; Flagellin_N; 1.
 DR PRINTS: PR00207; FLAGELLIN.
 DR ProDom: PD000316; Flagellin_C; 1.
 KW Flagella; Complete proteome.
 FT INIT_MET 0
 FT CONFLICT 284 284 P -> L (IN REF. 2).
 SO SEQUENCE 497 AA; 51163 MW; 24B5419C21C7B4E8 CRC64;

alignment_scores:
 Quality: 131.00 Length: 584
 Ratio: 0.468 Gaps: 26
 Percent Similarity: 47.945 Percent Identity: 20.377

alignment_block:

US-09-049-696-20 x FLIC_ECOLI ..

Align seg 1/1 to: FLIC_ECOLI from: 1 to: 497

```

970 AGCATGGCGACTGGTAACCGCTCAATCGACTGAATCAAGCAGGCCAGCT 1019
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10 SerLeuIleThrGlnAsnAsnIleAsnLys...AsnGlnSerAlaLeuSe 25
1020 TTTCCTGCTGCAGACAGTTGACTGGCTGGCTGGCTGGCTGGCTGGCT 1069
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
25 rSerSerIleGluArgLeuSerSerGly..... 34
1070 TTGACAGTGTGCTGCCATGTACAAAGTGAATCATACAGATAACAGTGGC 1119
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
35 .....LeuArgIleAsnSerAla 40
1120 AGTCAGAGGACACACTGCCAAAAGATTACCTGCAGCAGCTTCAGGAGG 1169
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
41 LysAsp...AspAlaAlaGlyGln.....AlaIleAlaAsnArgPh 53
1170 GAGCTCCATCTCCAGCGGCTTCGATCGGCATTTACTGTGATTAGGAAGA 1219
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
53 eThrSerAsnIleLysGlyLeuThrGlnAla.....AlaA 65
1220 AATATCAACTGATGGATCTGAATTTGTCTGCTGACGGATGGGAAGAC 1269
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 rGlnAlaAsnAspGlyIleSerValAlaGlnThrThrGluGlyAlaLeu 81
1270 AACACTATAAGTCGGTCTTAAACGAGGTCAACAA..... 1305
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
82 SerGluIleAsnAsnLeuGlnArgValArgGluLeuThrValGlnAl 98
1306 .....AGTGGTGCCATCATCCACACAGTCGCTTTGGGGCCCTCTG 1345
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98 aThrThrGlyThrAsnSerGluSerAspLeuSerSerIleGlnAspGluI 115
1346 CAGCTCAAGAACTAGAGGAGCTGTCCAAATGACAGGAGTTTACAGACA 1395
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
115 leLysSerArgLeuAspGluIleAspArgValSerGly..... 127
1396 TATCTTCATCAAGTTCAGAACTATGGC.....CTCATTTGATGCTTT 1439
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
128 .....GlnThrGlnPheAsnGlyValAsnValLeuAlaLysAs 140
1440 TGGGGCCTTTTCATCAGGAATAGGAGTGTCTCTCAGCGCTCCATC.... 1485
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
140 nGlySerMetLysIleGlnValGlyAlaAsnAspAsnGlnThrIleThrI 157
1486 .....CAGCTTGAGAGTAAGGATTACCTCCAGACACGCCAG 1524
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
157 leAspLeuLysGlnIleAspAlaLysThrLeuGlyLeuAspGlyPheSer 173
1525 TGGATGATGCACAGTGCCTGGACAGC..... 1554
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
174 ValLysAsnAsnAspThrValThrThrSerAlaProValThrAlaPheG1 190
1555 .....ACGTGGGAAGGACACTTTGTTTCTTATCACCTGGACACGCC 1597
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

190 YAlaThrThrThrAsnAsnIleLysLeuThrGlyIleThrLeuSerThrG 207
1598 AGCTCCCAAAATCCTTCTCTGGGATCCAGTGGGACAGCAAGCAAGTGGC 1647
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
207 luAlaAlaThr.....AspThrGlyThrAsnProAlaSer 219
146 ..... 146
1648 TTT.....CTAGTGACAAAACACAAATGCGCTACCTCCAAATCCC 1691
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
220 IleGluGlyValThrThrAspAsnGlyAsnAspTyrTyrAlaLysIleTh 236
1692 AGCAATTGCTAAGTGGGCACCTTGGAATACAGTCTGCAAGCAAGCTCAC 1741
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 rGlyGlyAspAsnAspGlyLysTyrTyrAlaValThrValAlaAsnAspG 253
1742 AAACCTTGACCCCTGACTGTCACGTCGCCGTCGTCCTCAATGTACCTGCCT 1791
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
253 LyThrValThrMet...AlaThrGlyAlaThrAlaAsnAlaThrValThr 268
1792 CCAATTACAGTGACTTCCAAAACGCAACAGACACACCAAAATTCCCCAG 1841
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
269 AspAlaAsnThrThrLysAlaThr..... 276
1842 CCCTCTGCTAGTTTATGCAAAATATTCGCCAAGAGCGCTCCCAATTCTCA 1891
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
277 .....ThrIleThrSerGlyThrProVal...G 286
1892 GGGCAGTGTGCACAGCCCTGATTGAATCAGTGAATGGAANAACAGTTACC 1941
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
286 InIleAspAsnThrAlaGlySerAlaThrAlaAsnLeuGlyAlaValSer 302
1942 TTGGAAGTACTGGATAATGGAGCAGGTGCTGCTACTACTAAGGATGACGG 1991
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
303 LeuValLysLeuGlnAspSerLysGlyAsnAspThrAspThr..... 316
1992 TGTCTACTCAAGGTATTTCACAACTTATGACACCAATGGTAGA...TACA 2038
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
317 .....TyrAlaLeuLysAspThrAsnGlyAsnLeuTyrA 328
2039 GTGTAAAGTCGGGCTCTGGGAGGAGTTACGCGACGACGAGGAGAGTG 2088
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 laAlaAspValAsnGluThrThrGlyAlaValSerValLysThrIleThr 344
2089 ATACCCACAGCAGTGGAGCAGTCTACATACCTCGCTGGGATGAGAATGA 2138
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345 TyrThrAspSerSerGlyAlaAlaSer..... 353
2139 TGAATACAAATGGATCCACCAAGCCTGAATTAATGAAGATGATGTTTC 2188
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
354 .....SerProThrAlaValLysLeuGlyGlyAsp..... 364
2189 AACACAAGCAAGTGTGTTTCAGCAGAACATCTCGGGAGGCTCATTTGTG 2238
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
365 .....GlyLysThrGlu 368
2239 GGTCTGTGATGTCCTCAATGCTCCCAATACCTGATCTCTCATTAAT... 2334
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
369 ValValAspIle.....AspGlyLys 375
2269 ATACCCAGCCTGAAGCGGAAATTCACGGGGCAGTCTCATTAAT... 2334
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
375 sThrTyrAsp...SerAlaAspLeuAsnGlyGlyAsnLeuGlnThrGlyL 391
2335 .....CTGACTTGGACAGTCTCTGGGATGATTAT 2364
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
391 euThrAlaGlyGlyGluAlaLeuThrAlaValAlaAsnGlyLysThrThr 407
2365 GACCATGGAACAGCTCACAGTATATCATTCGAATAAGTACAGTATTCT 2414
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
408 AspProLeuLysAla.....LeuAspAspAlaIleAl 418
2415 TGATCTCAGACAAAGTTCAATGAATCTCTCAAGTGAATACTACTGCTC 2464
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
418 aserVal...AspLysPheArgSerSerLeuGlyAlaValGlnAsnArgL 434
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

RA 256 leAspThrThrAspLysThrProLeuAlaThrAlaGluAlaThrAlaIle 272
RA 1792 .....CCAAATACAGTGCCTCCAAAC 1814
RA 273 ArgGlyThrAlaThrIleThrHisAsnGlnIleAlaGluValThrLysGI 289
RA 1815 GAACAAGGACACACAGCAAAATCCCCAGCCCTCTGGTAGTTATGCAATA 1864
RA 289 uGlyValAspThrThrValAlaAlaGlnLeuAlaAla...AlaGlyV 305
RA 1865 TTCGCCAAGGAGCCTCCCAATTCAGGCCAGTGTCAAGCCCTGATT 1914
RA 305 al...ThrGlyAlaAspLysAspAsnThrSerLeuValLysLeuSerPhe 320
RA 1915 GAATCAGTGAATGGAACAAACAGTTACCTTGGAACTACTGGATAATGGAGC 1964
RA 321 GluAspLysAsnGlyLysValle.....AspGlyGI 331
RA 1965 AGGTGCTGATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCACAA 2014
RA 331 yTyrAlaValLysMetGlyAspAsp.....PheTyrAlaAlaI 344
RA 2015 CTTATGAC.....ACGAAT 2028
RA 344 hrTyrAspGluLysGlnValGlnLeuLeuLeuAsnHisTyrThrAsp 360
RA 2029 GGTATACATACAGTGTAAAAGTCGGGCTCTG.....CGAGGAGTTAAGC 2072
RA 361 GlyAlaGlyValLeuGlnThrGlyAlaValLysPheGlyGlyAlaAsn... 376
RA 2073 ACCCAGCAGGAGTGATACCCAGCAGAGTGAGGACCTGTACATACCTG 2122
RA 377 GlyLysSerGluValValThrAlaThrValGlyLysThrTyrLeuAla 393
RA 2123 GCTGATTCAGATGATGAATACATGAATCCACCAAGACCTGAAT 2172
RA 393 erAspLeuAspLysHisAsnPheArgThrGlyGlyGluLeuLysGluVal 409
RA 2173 AATAAGGATGATGTTCAACAC 2193
RA 410 AsnThrAspLysThrGluAsn 416
seq_name: SwissProt_39:RPB1_DROME
seq_documentation_block:
ID RPB1_DROME STANDARD: PRT; 1887 AA.
AC P04052; 09VYX6;
DT 01-NOV-1986 (Rel. 03, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6).
GN RPII215 OR CG1534.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Nephroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89218930; PubMed=2496296;
RA Jolkerst R.S., Weeks J.R., Zehring W.A., Greenleaf A.L.;
RT "Analysis of the gene encoding the largest subunit of RNA polymerase
II in Drosophila.";
RL Mol. Gen. Genet. 215:266-275(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceinkner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,

```

```

Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
Baillet R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Besson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
Bokkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
Bujals K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
dePablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris D., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
Jatani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mout S.M., Moy M., Murphy B., Murphy L., Mazny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J., Yang Q.A.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [3]
RP SEQUENCE OF 1-27 FROM N.A.
RX MEDLINE=87089662; PubMed=3025586;
RA Searles L.L., Greenleaf A.L., Kemp W.E., Voelker R.A.;
RT "Sites of P element insertion and structures of P element deletions in
the 5' region of Drosophila melanogaster RpiI215.";
RL Mol. Cell. Biol. 6:3312-3319(1986).
RN [4]
RP SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=85282618; PubMed=2992806;
RA Biggs J., Searles L.L., Greenleaf A.L.;
RT "Structure of the eukaryotic transcription apparatus: features of the
gene for the largest subunit of Drosophila RNA polymerase II.";
FL Cell 42:611-621(1985).
RN [5]
RP SEQUENCE OF 1441-1887 FROM N.A.
RX MEDLINE=88094402; PubMed=3122024;
RA Allison L.A., Wong J.K.-C., Fitzpatrick V.D., Moyle M., Ingles C.J.;
RT "The C-terminal domain of the largest subunit of RNA polymerase II of
Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a
conserved structure with an essential function.";
RL Mol. Cell. Biol. 8:321-329(1988).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
CC RNA(N).
CC -1- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM TEN TO
CC TWELVE DIFFERENT POLYPEPTIDES. THE 215 KDA POLYPEPTIDE IS THE
CC LARGEST COMPONENT OF RNA POLYMERASE II.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC -1- THE PHOSPHORYLATION ACTIVATES POL2.
CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND TRNA GENES.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.

```


FT	1	21	POTENTIAL.
SIGNAL			CADHERIN-RELATED TUMOR SUPPRESSOR
CHAIN	22	4590	HOMOLOG.
FT			EXTRACELLULAR (POTENTIAL).
DOMAIN	22	4183	POTENTIAL.
TRANSMEM	4184	4204	CYTOPLASMIC (POTENTIAL).
FT			
DOMAIN	4205	4590	

CADHERIN 1:
CADHERIN 2

FT	DOMAIN	237	301	CADHERIN 3.
FT	DOMAIN	362	463	CADHERIN 4.
FT	DOMAIN	464	569	CADHERIN 5.
FT	DOMAIN	570	716	CADHERIN 6.
FT	DOMAIN	717	822	CADHERIN 7.
FT	DOMAIN	823	927	CADHERIN 8.
FT	DOMAIN	928	1034	CADHERIN 9.
FT	DOMAIN	1035	1138	CADHERIN 10.
FT	DOMAIN	1139	1245	CADHERIN 11.
FT	DOMAIN	1246	1345	CADHERIN 12.
FT	DOMAIN	1346	1456	CADHERIN 13.
FT	DOMAIN	1457	1562	CADHERIN 14.
FT	DOMAIN	1563	1670	CADHERIN 15.
FT	DOMAIN	1671	1769	CADHERIN 16.
FT	DOMAIN	1770	1882	CADHERIN 17.
FT	DOMAIN	1883	1982	CADHERIN 18.
FT	DOMAIN	1983	2084	CADHERIN 19.
FT	DOMAIN	2085	2185	CADHERIN 20.
FT	DOMAIN	2186	2286	CADHERIN 21.
FT	DOMAIN	2287	2393	CADHERIN 22.
FT	DOMAIN	2394	2495	CADHERIN 23.
FT	DOMAIN	2496	2599	CADHERIN 24.
FT	DOMAIN	2600	2705	CADHERIN 25.
FT	DOMAIN	2706	2811	CADHERIN 26.
FT	DOMAIN	2812	2920	CADHERIN 27.
FT	DOMAIN	2921	3024	CADHERIN 28.
FT	DOMAIN	3025	3127	CADHERIN 29.
FT	DOMAIN	3128	3232	CADHERIN 30.
FT	DOMAIN	3233	3337	CADHERIN 31.
FT	DOMAIN	3338	3442	CADHERIN 32.
FT	DOMAIN	3443	3546	CADHERIN 33.

DOMA1N	347	CADHERIN 34.
FT	3827	EGF-LIKE 1.
DOMA1N	3836	EGF-LIKE 2.
FT	4004	EGF-LIKE 3.
DOMA1N	4013	EGF-LIKE 4.
FT	4051	EGF-LIKE 5.
DOMA1N	4052	N-LINKED (GLCNAC).
FT	4089	
DOMA1N	4090	
FT	4126	
DOMA1N	4127	
FT	4163	
GARBOHYD	440	
FT	460	
FT	463	
FT	463	

FT	CARBOHYD	333	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	660	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	740	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	791	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	998	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	1426	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	GABOBYD	1551	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	1751	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	1867	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	1905	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARSOHYD	1943	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CAREOBYD	1994	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	2328	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	2467	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	3326	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	3424	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	3446	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	3615	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	3642	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	3718	N-LINKED (GLCNAC..)	(POTENTIAL)
FT	CARBOHYD	4154	N-LINKED (GLCNAC..)	(POTENTIAL)
SQ	SEQUENCE	4590 AA; 506273 MW; 004483CCDD000E60A7 CRC64;		

Length: 1195

```

Ratio: 0.248          Gaps: 62
Percent Similarity: 43.766      Percent Identity: 19.247

alignment_block:
US-09-045-696-20 x FATH_HUMAN  ..

Align seg 1/1 to: FATH_HUMAN from: 1 to: 4590

55  TTGATTCTTCACCTTCTAGAAGGGCCCTGAGT.....AATTCACCTCAT 98
   ||| : : : : : ||| : : : : : : : : : : : : : : : : : :
2648 LeuGluLeuAsnLysLeuSerGlyValIleThrThrLysGluSerLeuI 2664
   99  TCAGCTGAACACAAATGCTATGAAGGCGATTGCTGCTCAATCGACCCCA 148
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : :
2664 eGlyLeuGluAsnGluPheThrPhePheValargAlaValAspAsnG 2681
   149  ATGTGCCCAAGAGATGAACACTCATT...CAACAATAAAGGACATGGTG 195
      : : ||| : : : : : ||| : : : : : : : : : : : : : :
2681 lySerProSerLysGluSerValValLeuValTyValLysIleLeuPro 2697
   196  ACCCAGGACATCTCTGTATCTGTTTGAAGCTACAGAAACGCGATTTATT 245
      : : : : : ||| : : : : : : : : : : : : : : ||| :
2698 ProGluMetGlnLeuProLysPhe.....SerGluProPheTyThr 2711
   246  CAAAAATGTGCCATTTGATCTCCTGAAACATGGAAGCAAAAGCGTGACT 295
      : : ||| : : : : : ||| : : : : : ||| : : : : :
2711 rPheThrValSerGluAspValProVal.....GlyThrGluIleAspL 2726
   296  ATGTGAGACCAAACTGTGAGACCTCAAAAATGTGCTGATGTT..... 336
      : : ||| : : : : : ||| : : : : : ||| : : : : :
2726 euIleArgAlaGlu.....HisSerGlyThrValLeuTyrSer 2738
   337  CTGGTGTGAGCTACTCTCCAGTAAT...GATGAACCCCTACACT... 381
      ||||| : : : ||||| : : : ||||| : : : ||||| : : :
2739 LeuValLysGlyAsnThrProGluSerAsnAArgSpGluSerPheValII 2755
   382  ...GAGCAGATGGCACTGTGGAGAGAGGCTGAAAGGATCCACCTCA 427
      : : ||| : : : : : ||| : : : : : ||| : : : : :
2755 eAspArgGlnSerGlyArgLeuLysLeuGluLysSerLeuAspHisGlu 2772
   428  CTCCTGATTCATTGCAGGAAAAAGCTTAGCTGAATATGGACCACCAAGGT 477
      || : : : : : ||| : : : : : ||| : : : : : : :
2772 hrThrLysTyrGlnPheSerIleLeuAlaArgCysThrGlnAspAsp 2788
   478  AGGCATTTGTCATGAGTGGGCTCATCTACAGATGGGAGTATTTGACGA 527
      : : : ||| : : : : : ||| : : : : : ||| : : :
2789 HisGluMetValAlaSer...ValaspValSerIleGlnValLysAsp 2804
   528  GTACAATAATGAT.....GAGAAATCTTACT 553
      : : : ||| : : : ||| : : : ||| : : : ||| : :
2804 aAsnAspAsnSerProValPheGluSerSerProTyrgluAlaPheIle 2821
   554  TATCCCAAT.....GGAAGATACACAGCAGTAAGATGTTTCACGA 591
      : : : ||| : : : ||| : : : ||| : : : ||| : :
2821 aIleGluAsnLeuProGlyGlySerArgValIleGlnIleArgAlaSerAsp 2837
   592  GGTATTACTGGTACAATGTAGTAAGAAGTGTCCAGGGAGGCAGCTGTTA 641
      : : : : : ||| : : : : : ||| : : : : : ||| : :
2838 AlaAspSerGlyThrAsn.....GlyGlnValMetTy 2848
   642  CACCAAAAGATGCACATTCAATAAAGTAAACAGGACTCTATGAAAAAGGAT 691
      : : : : : ||| : : : : : ||| : : : : : ||| : :
2848 rSer..... 2849
   692  GTGAGTTTGTTCCTCAATCCGCCAGACGGAGAGGCTTCTATAATGTTT 741
      ||| : : : : : ||| : : : : : ||| : : : : : ||| :
2850 .....LeuAspGlnSerGlnSerValGluValIleGluSerPhe 2862
   742  GCACACATGTTGAT.....TCTATAGTTGAATTCGTGAC 776
      ||| : : : : : ||| : : : : : ||| : : : : : ||| :
2863 AlaIleAsnMetGluThrGlyTrpIleThrThrLeuLysGluLeuAspHi 2879
   777  AGACAAACACCAAC.....AAAGAAGCTCCAAACAGCAAAATC 817

```



```

3156 aGlyLeuAsnArgLysIleLeuTyrSerLeuIleAspSerAlaAspGlyG 3173
1460 ATGAGCTGTCTCTCAGCGCTCC.....ATCCAGCTTGAG..... 1494
:: : : : : : : : : : : : : : : : : : : : : : : : : : :
3173 InPheSerIleAsnGluLeuSerGlyIleIleGlnLeuGluLysProLeu 3189
1495 .....AGTAA 1499
3190 AspArgGluLeuGlnAlaValTyrThrLeuSerLeuLysAlaValAspGI 3206
1500 GGGATTAAACCTCCAGAACAGCAGCTGGATGAATGGCACAGTGTGGTGG 1549
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3206 nGlyLeuProArgLeuThr.....AlaThrGlyThrValIleValS 3221
1550 AC..... 1551
3221 erValLeuAspIleAsnAspAsnProProValPheGluTyrArgGluTyr 3237
1552 ..AGCACGTGGGAAAGGACACTTTGTCTTATCATCCTGGCAACGCA 1598
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3238 GlyAlaThrValSerGluAspIleLeuVal.....GlyThrGI 3250
1599 GCCTCCCCAAATCTCTCTGGGATGCC..... 1626
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3250 uValLeuGlnValTyrAlaAlaSerArgAspIleGluAlaAsnAlaGluI 3267
1627 .....AGTGGACAGAACAGGTGGCTTTGTAGTGGAC 1659
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3267 leThrTyrSerIleIleSerGlyAsnGluHisGlyLysPheSerIleAsp 3283
1660 AAAACACCAAAATGGCC..... 1677
3284 SerLysThrGlyAlaValPheIleIleGluAsnLeuAspTyrGluSerSe 3300
1678 .....TACCTCCAAATC.....CCAGGCA 1696
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3300 rHisGluTyrTyrLeuThrValGluAlaThrAspGlyGlyThrProSerL 3317
1697 TTCTTAGGTGGCACTTGGAAATACAGTCTG..... 1728
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3317 euSerAspAlaIaThrValAsnValAsnValThrAspIleAsnAspAsn 3333
1729 ..CAAGCAAGCTCACAAACCTTGACCTGACCTGACGTCGCGTCCGCTC 1775
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3334 ThrProValPheSerGlnAspThrTyrThrValIleSerGluAspAI 3350
1776 CAATGTACCTCCCTCCCAATTACAGTACTTCCAAACAGCAACAGGACA 1825
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3350 aValLeuGluGlnSerValIleThrValMetAla.....AspA 3363
1826 CAGCAAAATCCCGACGCTCTGGTAGTTATGCAATATT..... 1866
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3363 spAlaAspGlyProSerAsnSerHisIleHisTyrSerIleAspGlyI 3379
1867 CGCAAGGAGCCCTCC.....CCAATTCTC...AGGCCAGTGT 1901
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3380 AsnGlnGlySerSerPheThrIleAspProValArgGlyGluValLysVa 3396
1902 CACAGCCCTGATT.....GAATCACTGAATGGAACAGTTTACCTTGG 1945
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3396 lThrLysLeuLeuAspArgGluThrIleSerGlyTyrThrLeuThrValG 3413
1946 ACTACTGGATATGAGCAGGT..... 1968
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3413 lAlaSerAspAsnGlySerProProArgValAsnThrThrThrValAsn 3429
1969 .....GCTCATCTACTAAGGATGACGGTGTCTACTCAAGGTATT 2009
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3430 IleAspValSerAspValAsnAspAlaProValPheSerArg..... 3444
2010 CACAACCTTATGACACCAATGGTAGATACAGTGTAAAAGTGGCGGCTCG 2059
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3445 .....GlyAsnTyrSerValIleIleGlnGluAsnL 3455

```

```

2060 GAGGAGTTTAAC....GCAGCCAGACGAGAGTGTATACCCCGACAGAGTGA 2106
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3455 ysProValGlyPheSerValLeuGlnLeuValThrAspGluAspSer 3471
2107 GCACTGTACATACCTGGCTGGATT.....GAGAATGATGA 2141
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3472 SerHisAsnGlyProProPhePheThrIleValThrGlyAsnAspGI 3488
2142 A.....ATACAATGGAATCCACCAAGACCTGAAATTAATAAGGATGATG 2185
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3488 uLysAlaPheGluValAsnProGlnGlyValLeuLeuThrSerSerAlaI 3505
2186 TTCAACACAAAGCAA.....GTGTGTTTCAGCAGAACATCCTCG 2223
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3505 leLysArgLysGluLysAspHisTyrLeuLeuGlnValLysValAlaAsp 3521
2224 GGAGGCTCATTTTGTGGCTTCTGATGTCCCAAAATGCTCCCATTA..... 2265
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3522 AsnGlyLysProGlnLeuSerSerLeuThrTyrIleAspIleArgValII 3538
2266 ...CCTGATCTCTCCCACTGGCCAAATACCCGACCTGAAG..... 2304
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3538 eGluGluSerIleTyrProProAla...IleLeuProLeuGluIlePheI 3554
2305 .....GCGGAAATTCACGGGGGCGAGTCTCATTAATCTGACTGG 2343
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3554 leThrSerSerGlyGluGluTyrSerGlyGlyValIleGlyLysIleHis 3570
2344 ACAGCTCCTGGGATGATTATGACCATGGAACA..... 2376
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3571 AlaThrAspGlnAspValTyrAspThrLeuThrTyrSerLeuAspProGI 3587
2377 .....GCTC 2380
3587 nMetAspAsnLeuPheSerValSerThrGlyGlyLysLeuIleAlaH 3604
2381 ACAAG.....TATATCATTCGAATAAGT..... 2403
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3604 lSylLysLeuAspIleGlyGlnTyrLeuLeuAsnValSerValThrAsp 3620
2404 .....ACAAGTATTCTTGATCTC..... 2421
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3621 GlyLysPheThrThrValAlaAspIleThrValHisIleArgGlnValTh 3637
2422 .AGACAAAGTTCAATCAATCTCTTCAAGTGAATACTACTGCTCTCATCC 2470
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3637 rGlnGluMetLeuAsnHisThrIleAlaIleArgPheAlaAsnLeuThrp 3654
2471 CAAGGAAGCCCAACTCTGAG.....GAAGTCTTTTGTGTTTAAACCAGAA 2514
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3654 roGluGluPheValGlyAspTyrTrpArgAsnPheGlnArgAlaLeuArg 3670
2515 ACATTACTTTTGAAATGGCAGAGATCTT...TTTATTCTTATTTCAGGC 2561
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3671 AsnIleLeuGlyValArgArgAsnAspIleGlnIleValSerLeuGlnSe 3687
2562 TGTGTGAAGGTGATCTGAAATCAGAAATATCCAACATTGCACAGATAT 2611
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3687 r.....SerGluProHisProHisLeuAspValL 3697
2612 CTTTGTGTTTATTCTCCACAGACTCCGCCAGAGACACTAGTCTGTATGAA 2661
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3697 euLeuPheVal.....GluLysProGlySerAlaGln 3707
2662 ACCTCTGCTCTTGTCTCTAATATTCATATCAACAGCACCATTCTCT... 2706
: : : : : : : : : : : : : : : : : : : : : : : : : : :
3708 IleSerThrLysGlnLeuLeuHisLysIleAsnSerSerValThrAspII 3724
2707 .....GGCATTACACATTTTAAAAAATTATG..... 2730
3724 eGluGluIleIleGlyValArgIleLeuAsnValPheGlnLysLeuCysA 3741

```

2731TGAAGTGGATAGAGAA 2748
 |||||
 3741 laGlyLeuAspCysProTrpLysPheCysAspGlu 3752

seq_name: SwissProt_39:NEO1_MOUSE

seq_documentation_block:

ID NEO1_MOUSE STANDARD; PRT; 1493 AA.
 AC P97798; 2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NEOGENIN PRECURSOR.
 GN NEO1 OR NGN.

OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.

RC TISSUE=Brain;

RX MEDLINE-97407661; PubMed-9264410;

RA Keeling S.L., Gad J.M., Cooper H.M.;

RT "Mouse neogenin, a DCC-like molecule, has four splice variants and is expressed widely in the adult mouse and during embryogenesis.";

RL Oncogene 15:691-700(1997).

CC -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
 CC TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
 CC DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
 CC MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -!- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; 1 (SHOWN HERE), 2, 3, 4
 CC AND 5; ARE PRODUCED BY ALTERNATIVE SPLICING. THE EXPRESSION OF
 CC ISOFORMS 3, 4 AND 5 ARE DEVELOPMENTALLY REGULATED.

CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.

CC -!- DEVELOPMENTAL STAGE: EXPRESSED UBICUOUSLY THROUGHOUT THE MID TO
 CC LATE STAGES OF GESTATION AND IN ADULT TISSUES. STRONG EXPRESSION
 CC IS OBSERVED IN THE VENTRAL REGION OF THE VENTRICULAR ZONE OF THE
 CC E15.5 MOUSE NEURAL TUBE, AS WELL AS IN THE VENTRICULAR ZONES OF
 CC THE MESENCEPHALON AND RHOMBENCEPHALON. ISOFORMS 3 AND 4 ARE
 CC EXPRESSED AT HIGHER LEVEL COMPARED TO OTHER ISOFORMS BETWEEN E11.5
 CC AND E16.5.

CC -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.

CC -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
 CC TUMOR SUPPRESSOR PROTEIN DCC.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: Y09535; CAA70727.1; -

DR HSDP; P02751; ITTG.

DR MGD; MGI:1097159; Neol.

DR InterPro; IPR001777; FN_III.

DR InterPro; IPR003006; Ig_MHC.

DR InterPro; IPR003598; Ig_c2.

DR Pfam; PF00041; fn3; 6.

DR Pfam; PF00047; ig; 4.

DR PRINTS; PR00014; FNTYPEIII.

DR SMART; SM00060; FN3; 6.

DR SMART; SM00408; IgC2; 4.

DR Transmembrane; Immunoglobulin domain; Glycoprotein; Signal;

KW Alternative splicing.

FT SIGNAL 1 36 POTENTIAL.

FT CHAIN 37 1493 NEOGENIN.

FT DOMAIN 37 1136 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1137 1157 POTENTIAL.

FT DOMAIN 1158 1493 CYTOPLASMIC (POTENTIAL).

DOMAIN 78 147 IG-LIKE C2-TYPE DOMAIN 1.
 DOMAIN 177 239 IG-LIKE C2-TYPE DOMAIN 2.
 DOMAIN 274 338 IG-LIKE C2-TYPE DOMAIN 3.
 DOMAIN 366 428 IG-LIKE C2-TYPE DOMAIN 4.
 DOMAIN 467 564 FIBRONECTIN TYPE-III 1.
 DOMAIN 567 660 FIBRONECTIN TYPE-III 2.
 DOMAIN 661 760 FIBRONECTIN TYPE-III 3.
 DOMAIN 766 860 FIBRONECTIN TYPE-III 4.
 DOMAIN 881 981 FIBRONECTIN TYPE-III 5.
 DOMAIN 982 1083 FIBRONECTIN TYPE-III 6.
 DOMAIN 1149 1153 POLY-VAL.
 FT CARBOHYD 84 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 221 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 337 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 520 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 670 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 746 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 940 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPLIC 442 MISSING (IN ISOFORM 2).
 FT VARSPLIC 863 MISSING (IN ISOFORM 3).
 FT VARSPLIC 1086 MISSING (IN ISOFORM 4).
 FT VARSPLIC 1279 MISSING (IN ISOFORM 5).
 SQ SEQUENCE 1493 AA; 163159 MW; 441DE919D5E17C0E CRC64;

alignment_scores:

Quality: 129.00 Length: 742
 Ratio: 0.430 Gaps: 42
 Percent Similarity: 40.431 Percent Identity: 20.620

alignment_block:

US-09-049-696-20 x NEO1_MOUSE

Align seg 1/1 to: NEO1_MOUSE from: 1 to: 1493

898 CCACCAATCCACCTTCTCATTCGACATTCGACAAAGAAATGTGTG 947
 ||||| ||| ||| ||||| ||| ||||| |||||
 21 ProProProLeuLeuLeuLeuProLeu.....LeuLeuLe 34
 943 TTATGCTCTTGACAAATCTGAACATGGCGACTGGTAACGGCTCAATC 997
 ||| ||||| ||||| ||||| ||||| ||||| |||||
 34 uLeuGlyArgProAlaSerGlyAlaAlaThrLysSerGlyProArg 51
 998 GACTGAATCAAGCAGCGCAGCTT.....TTCTGTG 1026
 || ||||| ||||| ||||| ||||| ||||| ||||| |||||
 51 rGlnSerGlnGlyAlaSerValArgThrPheThrProPheTyrPheLeu 67
 1027 CTGCACAGACGTGAGCTGGGTCTCTGGTGGGATGGTGACATTTGACAG 1076
 .: ||||| |||
 68 ValGluProVal..... 71
 1077 TGCTGCCCATGTACAAAGTGAACATCATACAGATAAAGTGGCAGTGACA 1126
 .: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 71 71
 1127 GGGACACACTGCCCAAGAAATTACCTGCAGCAGCTTCAGAGGAGCGTC 1176
 .: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 72 ..AspThrLeuSerValArg.....GlySerSer 80
 1177 ATC.....TGACGGGGCTTCGATCGGCATTTACTGTGATTAGGAA 1217
 .: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 81 ValIleLeuAsnCysSerAlaTyrSerGluProSerProAsnIle..... 95
 1218 GAAATATCCAACTGATGATCTGAAATTTGCTGCTGACGGATGGGAAG 1267
 .: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 95 .GluTrpLysLysAspGlyThrPheLeuAsnLeuGluSerAspArg. 111
 1268 ACAACACTATAAGTGGTGTCTTTACAGAGGTCACAAAGTGGTCC.... 1314
 .: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 112ArgGlnLeuLeuProAspGlySerLeu 120
 1315ATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCA 1352

```

121 PheIleSerAsnValValHisSerLysHisAsnLysProAsp..... 134
1353 AGAACTAGAGGAGCTGTCCAAAATGACAGGAGGTTTACACATATGCTT 1402
135 ..:::.....GluGlyPheThrGlnCysValAlaIat 143
1403 CAGATCAAGTTTACAGCAATGCGCTCATGTATGCTTTTGGGGCCCTTCA 1452
143 hr.....ValAspAsnLeuGlyThrIleVal 151
1453 TCAGGAAATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGG 1502
152 Ser.....ArgThrAlaLysLeuThrValAlaGl 161
1503 ATTAACCTCCAGAACAGCCAG.....TGGATGAATG 1534
161 yLeuProArgPheThrSerGlnProGluProSerSerValTyrValGlyA 178
1535 GCACAGTGTGCTGGACACCGCTGGGAAAGGACACTTTGTTCTTATC 1584
178 snSerAlaIleLeuAsnCysGluValAsnAlaAspLeuValProPheVal 194
1585 ACCTGGACAACGAGCTCTCCCAATCCCTCTCTGGGAT..... 1623
195 ArgTrpGluClnAsnArgGlnProLeuLeuLeuAspAspArgIleVally 211
1624 ...CCCAGTGA.....CAGAAGCAAGTG 1645
211 sLeuProSerGlyThrLeuValIleSerAsnAlaThrGluGlyAspGlyG 228
1646 GCTTTT.....GTAGTGGACAAACACCAAAATGCGCTACCTCCAA 1686
228 lLeuThrArgCysIleValGlnSerGly..... 237
1687 ATCCAGGCAATGCTAAGTGTGGCAATGGAATACAGT..... 1725
238 .....GlyProProLysPheSerAspGluAlaGl 247
1726 .....CTGCAAGCAAGCTGACAACTTGACCTGACGTGTCACGT 1765
247 uLeuLysValLeuGlnAspProGluGluIleValAspLeuValPheLeuM 264
1766 CCGTGCCTCC.....AATGCTACCTGCGCT 1791
264 eArgProSerSerMetMetLysValThrGlyClnSerAlaValLeuPro 280
1792 CCAATTACAGTGACTTCCAAACGAACAGGACACCAAGCAAAATCCCCAG 1841
281 CysVal.....ValSerGlyLeuProAl 288
1842 CCTCTGGTGTATGCAAAATATGCGCAAGGAGCGCTCCCAATTTCTCA 1891
288 aProValValArgTrpMetLysAsnGlu..... 298
1892 GGCCAGTGTACAGCCCTGATGATGATCAATGAGTGAAGAAACAGTTACC 1941
299 .....ValLeuAspThrGluSerSerGlyArgLeuValLeu 310
1942 TTGGAACACTGGAATAATGAGCAGCT.....GCTGATGC 1976
311 Leu.....AlaGlyGlyCysLeuGluIleSerAspVa 321
1977 TACTAAGGATGACGGTGTCTACTCAGGATTTTACAACTTATGACACGA 2026
321 lThrGluAspAspAla.....GlyThrThrPheCysIleAlaAsp..A 335
2027 ATGGTAGATACACTGTAAGTCCGGCTCTGGGAGGACTTTACGCGACC 2076
335 snGlyAsnLysThrValGluAlaGlnAlaGluLeuThrVal..... 348
2077 AGACGGAGAGTGATACCC.....CAGCAGAGTGAGCAGCTGTAC... 2115

```

```

349 .....GlnValProProGlyPheLeuLysGlnProAlaAsnIleTyrAl 363
2116 .....ATACCTGGC..... 2124
363 aHisGluSerMetAspIleValPheGluCysGluValThrGlyLysProT 380
2125 .....TGGATTGAGATGATCAATACAAATGGAATCCACCA 2160
380 hrProThrValLysTrpValLysAsnGlyAspValValIleProSerAsp 396
2161 AGACCTGAAATTAATAAGGATGATGTTCAACACAAAGTGTGTTTCAG 2210
397 AsnPheLysIleValLysGlu.....HisAsnLeuGlnValLeuGlyLe 411
2211 CAGAACATCCTCGGAGGCTCATTT.....GTGGCTTCTGATGCTC 2251
411 uValLysSerAspGluGlyPheThrGlnCysIleAlaGluAsnAspValG 428
2252 CAAATGCT.....CCC 2262
428 lYAsnAlaGlnAlaGlyAlaGlnLeuIleLeuGluHisAspValAla 444
2263 ATACCTGATCTCTCCACCTGGCCAAATCACC..... 2295
445 lIleProThrLeu...ProProThrSerLeuThrSerAlaThrThrAspHi 460
2296 .....GACCTGA 2302
460 sLeuAlaProAlaThrThrGlyProLeuProSerAlaProArgAspVal 477
2303 AGCGGAAATTCACGGGGCAGCTCTCATTAATCTGACTTGGACAGCTCCT 2352
477 aAlaSerLeuValSerThrArgPheIleLysLeuThrTrpArgThrPro 493
2353 GGGGATGATTATGACCATGGAACAGCTCACAGTATATCATTCGAATAAG 2402
494 AlaSerAsp...ProHisGlyAspAsnLeuThrTyrSerValPheThr 509
2403 TACAGTATTTGATCTCAGACAGCAAGTTCATGAATCT..... 2442
509 rLysGluGlyValAspArgGluArgValGluAsnThrSerGlnProGlyG 526
2443 ..CTCAAGTGAATACTACTGCTCTCATCCCAAGGAGCAACTCTGAG 2490
526 lMetGlnValThrIleGlnAsnLeuMetPro.....Ala 537
2491 GAAGTCTTTTGTAAACCCAGAAAACATTACTTTTGAATAATGGCACAGA 2540
538 ThrValTyrIlePheLys..... 543
2541 TCTTTTCATTGCTATTCAGGCTGTTGATAGGTCGATCTGAAATCAGAAA 2590
544 .....ValMetAlaGlnAsnLysHisGly...SerGlyGluS 555
2591 TATCCAACTTCACAGCATCTTTGTTTATTCCTCCACAGACTCCGCCA 2640
555 exSerAlaProLeuArgVal.....GluThrGlnPro 565
2641 GAGACACCTAGTCTGATGAAACGCTGCTGCTCTGCTTAATATTCATAT 2690
565 GluValGlnLeuPro.....GlyProAlaProAsnIleArgAl 578
2691 CAACAGCACCATTCCTGGCATTACACATTTTAAATAATTTATGTGGAAG... 2736
578 aThrAlaThrSerPro.....ThrSerIleThrValThrTrpGluThrP 593
2737 .....TGGATAGGAGAACTCCAG 2754
593 roLeuSerGlyAsnGlyGluIleGln 601

```

This Page Blank (uspto)

OM of: US-09-049-696-20 to: PIR_68.* out_format : pfs
Date: Mar 30, 2002 2:34 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters: -DEV=slp
-MODEL=frame1-n2p.model -USPTO=spool/US09049696/runat_28032002_145238_2031/app_query.fasta_1.12579
-DB=PIR_68 -OPMT=fastcan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09049696 -CGN1_1_367 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPYX -WAIT -THREADS=1

Search information block:
Query: US-09-049-696-20
Query length: 2983
Database: PIR_68.*
Database sequences: 219241
Database length: 76174552
Search time (sec): 470.790000

score_list:	Strd	Orig	ZScore	EScore	Len	Documentation
Sequence	+	171.50	199.77	0.0009	420	! hypothetical protein - Synchoc
pir2:S76691	+	153.50	165.31	0.0193	1672	! polymorphic membrane protein B
pir2:C81675	+	153.00	172.88	0.0177	689	! probable retroelement pol poly
pir2:F84811	+	152.00	159.22	0.0265	2660	! probable invasin Z3135 (impor
pir2:E85822	+	150.50	167.12	0.0276	928	! probable lipoprotein Cj1678 (in
pir2:C81265	+	145.50	145.12	0.0818	5188	! probable RTX family exoprotein
pir2:B85547	+	143.50	150.14	0.0965	2340	! cell surface antigen (scs3) RH
pir2:B71704	+	140.00	153.96	0.1415	978	! probable lipoprotein Cj0629 (in
pir2:D81411	+	138.50	147.99	0.1938	1536	! high-molecular-weight surface
pir2:A43855	+	138.00	154.85	0.1806	683	! hypothetical protein F3512.4b
pir2:T21810	+	136.50	153.00	0.2280	686	! hypothetical protein F3512.4a
pir2:T21808	+	136.50	146.03	0.2617	1461	! hypothetical protein 20609 (in
pir2:A85547	+	136.00	152.24	0.2471	698	! hypothetical protein MTH911 - M
pir2:A69222	+	136.00	143.26	0.2952	1849	! Iga-specific metalloendopeptid
pir2:C41859	+	135.50	148.83	0.2823	946	! inter-alpha-trypsin inhibitor
pir2:JC5575	+	135.50	142.35	0.3209	1912	! protein-tyrosine-phosphatase
pir2:A56178	+	135.50	143.10	0.3276	1651	! transmembrane receptor protein
pir2:T14160	+	134.50	147.54	0.3301	955	! probable membrane protein b1202
pir2:G64866	+	134.00	145.04	0.3703	1173	! prockr2 - chicken (fragment)
pir2:I50620	+	134.00	116.15	0.6555	26926	! titin, cardiac muscle [valida
pir1:I38344	+	133.50	146.39	0.3849	949	! hypothetical protein f949 - Esc
pir2:T09658	+	133.50	141.75	0.4219	1569	! hypothetical protein b2647 - E
pir2:A65044	+	133.50	144.95	0.4228	1039	! hypothetical protein T22A6.30
pir2:T09883	+	132.50	142.01	0.4785	1339	! 1,3-beta-D-glucan-binding prot
pir2:A55301	+	132.00	144.88	0.4826	918	! hypothetical protein - Acinetob
pir2:T02759	+	132.00	143.50	0.4960	1067	! probable extracellular nucleas
pir2:D75625	+	131.50	143.72	0.5273	976	! polymorphic membrane protein E
pir2:F81722	+	131.50	138.38	0.5860	1742	! projection - fruit fly (Drosoph
pir2:S24600	+	131.00	149.31	0.5040	498	! flagellin - Escherichia coli (s
pir1:FLEC	+	131.00	141.93	0.5832	1109	! surface-array protein homolog
pir2:A56143	+	130.50	148.58	0.5459	505	! flagellin - Salmonella muenche
pir2:S09638	+	130.50	139.83	0.6491	1305	! 150K mating aggregate formati
pir2:H41662	+	130.50	129.16	0.8015	4152	! filamentous hemagglutinin 1 -
pir2:T31102	+	130.00	138.31	0.7141	1441	! hypothetical protein yoiC (impr
pir2:B86807	+	130.00	135.79	0.7506	1894	! protein-tyrosine-phosphatase
pir2:C54689	+	130.00	134.28	0.7734	2232	! hypothetical protein K06A9.1a
pir2:T34434	+	129.50	142.38	0.7035	868	! CD22 homolog/B lymphocyte-rest
pir2:A46512	+	129.50	140.17	0.7350	1104	! probable membrane protein ymr3
pir2:S59310	+	129.50	135.18	0.8111	1896	! DNA-directed RNA polymerase (E
pir1:RNFF2L	+	129.00	137.24	0.8315	1421	! hypothetical protein At2g38440
pir2:T02501	+	129.00	136.57	0.8426	1528	! hypothetical protein ypiA (impr

pir2:G71518 + 128.50 134.71 0.9333 1751 ! hypothetical protein pmpB -
pir2:A83412 + 128.50 131.54 0.9936 2468 ! hypothetical protein PA1874
seq_name: pir2:S76691
seq_documentation_block:
hypothetical protein - Synchocystis sp. (strain PCC 6803)
C:Species: Synchocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S76691
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyaajima,
O.; K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocys
S.
A:Reference number: S74322; MUID:97061201
A:Accession: S76691
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-420 <KAN>
A:Cross:references: EMBL:D64004; GB:AB001339; NID:g1001701; PIDN:BAAL0635.1; PID:g101
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
alignment_scores:
Quality: 171.50 Length: 455
Ratio: 0.776 Gaps: 22
Percent Similarity: 48.571 Percent Identity: 23.956
alignment_block:
US-09-049-696-20 x S76691 ..
Align seg 1/1 to: S76691 from: 1 to: 420
775 ACAGAACAAACACACACAAAGAGCTCCAAACAGCAAAATCAAAAATG 824
10 SerAspGlnAsnLeuAspAlaGlyAlaProThrSerGlnArgGln..... 24
825 CAATCTCCGAAGCAGTGGGAAGTGTGATCGCTGATTCGTGAGGACTTTAAGA 874
25LeuArgIleAla.....ValalaAlaLysAlaAspHisAspA 38
875 AAACCACTCTATGACAAACACACGCCACCAATCCACCTTCTCATTTGCTG 924
38 rgArgLeuProLeuasn.....LeuCysLeuValLeuAspHisSerGlySerMe 54
925 CAGATTGGACAAAGAATTGTGTGTTTGTAGCTTGTACAAATCTCGAAGCAT 974
44LeuCysLeuValLeuAspHisSerGlySerMe 54
975 GCGCACTGGTAACCGCTCAATCGACTGAATCAAGCAGCGCCAGCTTTTCC 1024
54 t...AspGlyGlnProLeuGluThrValLysSerAlaAla...LeuGlyL 69
1025 TCGTCGACAGACTTGAGTGGGTCTCGGTGGGTGGGTGGGTGGGTGGGTGGT 1074
69 euIleAspArgLeuGluGluAspArgLeuSerValIleAlaPheAsp 85
1075 AGTGTGCGCCATGTACAAAGTGAACATACACAGATAACAGTGGCAGTGA 1124
86 HisArgAlaLysIleValIleGluasnGlnValArgasnGlyAlaAl 102
1125 CAGGGACACATCGCCAAAAGATTACTCGCAGCAGCTTTCAGGAGGACGT 1174
102 alleAlaLysAlaIleGluArgLeu.....LysAlaGluGlyGlyThra 117
1175 CCATCTGACGGGTTCGATCGGATTTACTGTGATTTAGGAAGAATAT 1224
117 latleAspGluGlyLeuLysLeuGlyIleGlnGluAlaLysGlyLys 133
1225 CCNACTGTGGATCTGAATTTGCTGCTGACCGATGGGGAACACACAC 1274

134 GluAspArgValSerHisIlePheLeuLeuThrAspGlyGluAsnGlu... 149
1275 TATAAGTGGTCTTTACGAGGTCACAAAGTGGTCCATC..... 1317
150 ...HisGlyAspAsnAspArgCysLeuLysLeuGlyThrValAlaSerA 165
1318ATCCACACAGCTCGCTTTGGGGCCCTCTGCAGCTCAA 1353
165 sPtyrLysLeuThrValHisThrLeuGlyPheGlyAspHisTrpAsnGln 181
1354 GAA...CTAGAGAGCTCTCAAAATGACAGGAGTTTACAGACATATGC 1400
182 AspValLeuGluAlaIleAlaSerAlaGlnGlySerLeuSerTrpI 198
1401 TTCAGAT.....GTCCTCAGCGCTCCATCCAGCT 1490
198 eGluAsnProSerGluAlaLeuHisThrPheArgGlnLeuPheGlnArgm 215
1412 TTCAGAACAAATGGCTCATATGATGCTTTGGGGCCCTTTTCATCAGGAAT 1461
215 etSerAsnValGlyLeuThrAsnAlaHisLeuLeuGluLeuAlaPro 231
1462 GGAGCT.....GTCCTCAGCGCTCCATCCAGCT 1490
232 GlnAlaHisLeuAlaIleValLysProValAlaGln.....ValSerPr 246
1491 TGAGAGTAAGGATTAAACCTCCAGACAGCCAGTGGATGAATGCCACAG 1540
246 oGluThrMetAspLeuThrValGlnAsnGlnGlyAlaIleGluGluVala 263
1541 TGATCGTGACAGACCGCTGGGAAGGACACTTTGTTCTTATCACCCTGG 1590
263 rgLeuGlyAspLeuMetThrAspGlnGluArgValLeuLeuLeuAsn... 278
1591 ACAACCCAGCTCCCAAAATCTCTCTGGGATCCAGTGGAGACAAGCA 1640
278 278
1641 AGTGCGCTTTGTAGTGGACAAAACACCAAAATGGCTACCTCCAA... 1686
279LeuTyrLeuAspGlnL 284
1687 ..ATCCAGGCACTGTAAAGTTGGC.....ACTTGGAAATACAGTCTG 1728
284 euleProGlyGlnHisValIleGlyGlnValGlnIleargTyrAspasp 300
1729 CAAGCAAGCTCACAAAACC.....TTGACCT 1754
301 ProAlaSerGlyGlnThrAsnLeuLeuSerAspProLeuProLeuThrI 317
1755 GACTGTGCTCCCGTCCGTCACATCTACCTCCCTCCCAATTACAGTGA 1804
317 eGlnValGlnThrGlnThrProSerThrAspValGlnValGlnGluS 334
1805 CTTCCAAAGGACAAAGGACACACAGCAAAATTTCCCGCCCTCTGGTA... 1851
334 erIleLeuThr.....LeuAlaLysTyrArgGlnThrGlnIleAla 347
1852GTTATGCAATATTCGCCAGGAGCCCTCCCAATTTCTCAG 1892
348 GluThrLysLeuLysAlaGlyAspArgGlnGlyAlaAlaThrMetLeuG 364
1893 G...GCCAGTGTCAACCCCTG...ATTGAATCAGTGAATGAAAAACAG 1936
364 nThrAlaAlaLysThrAlaLeuGlnMetGlyAspLysAsnGlyAlaThrI 381
1937 TTACCTTTGGAACTACTGGATAATGA.....GCAGGTGCTGAT 1974
381 ie.....LeuGlnThrAsnAlaThrArgLeuGlnSerGlyGluAsp 394
1975 CTACTAAGGATGAC 1989
395 LeuSerGluGlyAsp 399

seq_name: pir2:C81675

seq_documentation_block:
polymorphic membrane protein B/C family TC0694 [imported] - Chlamydia muridarum (stra
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: C81675
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Whinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: R81500; MUID:20150255
A:Accession: C81675
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1672 <TET>
A:Cross-references: GB:AE002338; GB:AE002160; NID:g7190724; PIDN:AAF39510.1; PID:g719
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0694

alignment_scores:

Quality: 153.50 Length: 802
Ratio: 0.468 Gaps: 34
Percent Similarity: 40.898 Percent Identity: 18.080

alignment_block:

US-09-049-696-20 x C81675 ..

Align seg 1/1 to: C81675 from: 1 to: 1672

547 TTCTACTTATCCAAATGCAAGTAACAGCAAGTAAGTGTTCACGAGTAT 596
122 PheTyrAsnGluHisSerGlyIleLeuSerPheMetAlaArgSerGlyVa 138
597 TACTGGT.....ACAAATGTAGTAAAGAAAGTGTCTCAGGGAGGCA 634
138 iGluGlySerLeuThrLeuSerAsnIleLysMetThrGlyAspGlyGlyA 155
635 GCTGTTTACACCAA...AGATGCACATTCATTAAGTAACAGGACTCTAT 681
155 laIleTyrSerGlnGlyGluLeuLeuPheThrAspLeuThrGlyLeuThr 171
682 CAAAAGGATGTGAGTTGTCTCCAAATCCCGCAGCAGCAGGAGAGGCTTC 731
172 IleGlnGly.....AsnLeuSerGlnLeuSerGlyGlyG 183
732 TATATGTTTCCACACATGTTGATTCTATAGTTGAATTCGTACAGAAC 781
183 yIlePheGlyGly.....SerThrIleSerPhe..... 192
782 AAACCCACAAAGAGAGCTCCAAACAAAGCAAAATCAAAATCAAAATCTC 831
193SerGlyIleAsnGlnAlaThrPheSer 201
832 CGAACACATGGGAAGTATCCGTGATTCCTGAGGACTTTAAGAAAACAC 881
202 SerAsnThrAlaGluValValProGlu.....GluThrTh 213
882 TCCT.....ATGACAAACAGCCACCAATCCCA 910
213 rProAsnProAsnProGlyThrGlnThrThrSerGlnProSerProT 230
911 CCTTCTCATTTGTCAGATTGGACAAAGAATTTGTTTAGTCTCTGAC 960
230 hr.....SerLysValGlnSerLeuPheThrTyr 239
961 AAATCTGGAAGCATCGGACTGGTAAACCCCTCAATCGACTCAATCAAGC 1010
240 SerSerSerThrGlnAlaAsnGlyAsn..... 248

1011 AGCCAGCTTTTCTGCTGCACAGAGTTGAGCTGGGCTCCTGGGTGGGA 1060
248
1061 TGGTGACATTTGCACAGTGTGCTGCCCATGTACAAAAGTGAACATCATACAGATA 1110
248
1111 AACAGTGGCAGTGACAGGGACACATCCGCCAAAAGATTACCTGCAGCAGC 1160
249GlyAlaAspSerGlnThrProSerHisLys.....ProG1 260
1161 TTCAGAGGGAGCTCCATCTGCAGGGGCTTCGATCGCATTTACTGTGA 1210
260 ySerGlyGlyAlaIleTyrAlaIleGly.....AspLeuThrIle. 273
1211 TTAGGAAGAAATATCCAATGATGATCTGAAATGTGTGCTGCAGCGAT 1260
274SerAspSerGlnGluIleValPhe..... 281
1261 GGGGAAGACAACTATAGTGGGTCTTAACGAGGTCAAAAGTGG 1310
282SerValAsnLysAlaSerLysAspG1 290
1311 TCCCATCATCCACAGCTGCTTTGGGCCCTCTGCAGCTCAAGACTAG 1360
290 yGlyAlaIle..... 293
1361 AGGAGCTGTCCAAAATGACAGGAGTTTACAGACATATGCTTCAGATCAA 1410
294PheAlaGluLysAsnValSerPheGluAsnIleThrThrLeuLys 308
1411 GTTCAGACAATGCGCTCATGATGCTTTGGGGCCCTTTCATCAGGAAA 1460
309 ValGlnAsnAsn.....GlyAlaGluGluLysGlyG1 319
1461 TGGAGTGTCTCTCAGCGCTCCATCAGCTTCAGAGTAAGGGATTAAACC 1510
319 yGlyIleTyrAlaSerGlyAspLeuSerIleGlnSerLysGlnSerL 336
1511 TCCAGAACAGCCAGTCGATGAATGGCACAGTGTGTCGACAGCACCGTG 1560
336 euPheAsnSer..... 339
1561 GGAAGGACACTTTGTTCTTATCACCCTGGACACGACCCCTCCCAAT 1610
339
1611 CCTTCTCTGGATCCAGTGGACAGCAAGTGGCTTTGTAGTGACA 1660
340AsnThrSerLysGlnGlyGlyAlaLeuTyrIleGluG 353
1661 AAACACCAAAATGGCTACCTCCAATCCAGGCAATGCTAAGGTGGC 1710
353 lyAsnValAspPheLysAspLeuGluIleArgIle..... 365
1711 ACTTGGAAATACAGCTGCAAGCAAGCTCACAACCTTCACCTGACTGT 1760
366LysTyrAsnLysSerGlyThrPheGluThrLysLysValThrLe 380
1761 CAGCTCCCGTGGCTCAATGCTACCTGCTCCAATTCAGTGACTTCCA 1810
380 uSerLeuProGluAlaGlnThrAsnLysSerSerValThrAlaIleSerG 397
1811 AACGACCAAGGACACCAAGCAATCCCGAGCCCTCTGTA..... 1851
397 InSerGlyProAsnThrThrProThrProThrProValThrAlaLys 413
1852GTTTATGCAATATTCCCAAGGAGCTCCCAATCTCTCAG 1892
414 GlyGlyGlyLeuTyrThrGluLysAsnLeuSerIleSer..... 426
1893 GGCCAGTGTACAGCCCTGATTGAATCAGTGAATGGAAAAACAGTTACCT 1942

427AsnIleThrGlyIleIleGluIleThrAsnAsnLysAlaThr..... 440
1943 TGGAACTACTGTAATGGAGCAGGTGCT..... 1971
441AspValGlyGlyGlyAlaTyrValLysGlyThrLeuThr 453
1972GATGCTACTAA 1982
454 CysLysAspSerHisArgLeuGlnPheGlnLysAsnSerSerGluLysLys 470
1983 GGATCAGCGTGTCTACTCAAGTATTTCACAACCTATGACACAGCAATGGTA 2032
470 sclyGlyGlyLeuTyr.....ThrGluAspThr..... 479
2033 CATACAGTGTAAAGTGGCGGCTCTGGGAGAGTTAAACGACGCCACGACGG 2082
480IleThrLeuSerAsnLeuThrGlyLysThrLeuPheGlnGlu 493
2083 AGAGTGATACCCAGCAGAGTGGAGCTGTACATACCTGGCTGGATTGA 2132
494 AsnThrAlaLysGluGluGlyGlyLeuTyrIleGlnGly..... 507
2135 GAATGATGAATACAATGGAATCCACCAAGACCTGNAATTATTAAGGATG 2182
508AspA 509
2183 AT.....GTTCAACACAAAGTGTGTTTC..... 2208
509 spLysThrLeuThrMetThrGlyLeuAspSerPheCysLeuIleAspAsn 525
2205ACGAGAACATCTCGGAGGCTCATTTGGCTTCGTGATGTCCCAA 2255
525 ThrSerAlaThrHisGlyGlyAlaTyrValThrLysGluIleSerG1 542
2256 TGCTCCCATACCTGATCTC.....TTCACCTGGCCAAATCACCGACC 2299
542 nThrThrThrSerAspValGluGluPhePro.....GlyIleThr..... 555
2300 TGAAGCGGAAATTCACGGGCGCAGTCTCATTAATCTGACTTTGGACAGCT 2349
556ProValHisGlyGluThrIleSerGlyAsnLysAlaThr 569
2350 CTGGGATGATTATGACCATGGAACAGCTCACAGTATATCATTCGAAT 2399
579 GlyGlySerGlyGlyValCysThrLysHisLeuValLeuSerAsnLe 586
2400 AAGTACAAGTATTCTGTATCTCAGACACAAGTTCAATGAATCTCTTCAAG 2449
586 uGlnThrIleSerIleSer..... 592
2450 TGAATACTACTCTCTCATCCCAAAGGAAGCAACTCTGAGGAAGTCTTT 2499
592
2500 TTGTTTAAACCAGAAACATTACTTTTGAATATGCCACAGATCTTTTCAT 2549
593GluAsnPheAlaSerGluAsnGlyGlyGly..... 602
2550 TGCTATTACAGCTGTGATAAGTCTGAAATCAGAAATATCAACA 2599
603 AlaCysThrCysProAspAsnPhe..... 610
2600 TTGCACGAGTATCTTTGTTTATTCCTCCACAGACTCCGCCAGACACCT 2649
611ProAlaProThrAlaSerThrPro 618
2650 AGTCTGTGAAACGCTCTCTCT.....TGTCCTAA 2681
619 SerThrAsnGlnThrAlaProLysAspLysAspPheLeuIleAs 635
2682 TATTCATATCAACAGCACCAT.....C 2704


```

326 ATGCTGATGTTCTGGTGTGAGTCTACTCTCCAGGTAATGATGAACC 375
|| : : : : : ||| : : : : : ||| : : : : : |||
216 sn.....LeuPheSerGlnHisThrLeuHisArgThrAspGlu... 229
376 TACACTGAGCAGATGGCAACTCTGGAGAGAGAGGTGAAGGATCCACCT 425
||| : : : : : ||| : : : : : ||| : : : : : |||
230 ...ArgThrGlnIleAsnAsn.....GlyLeuGlyTrpArg...HisPhe 242
426 CACTCTCTGATTTTCATTGCGAGGAAAAAGTTAGCTGAATATGGACCACAAG 475
||||| : : : : : ||| : : : : : ||| : : : : : |||
242 eThrProThrTrpMetSerGlyIleAsnPhe..... 252
476 GTAGGGCAATTGTCCATGATGGGCT.....CATCTACGATGGGAGTA 519
||| : : : : : ||| : : : : : ||| : : : : : |||
253 ....PhePheAspHisAspLeuSerArgTyrHisSerArgAlaGlyIle 267
520 TTTGACGAGTACAAATAATGATGAGAAATTTCTACTTATCCAAATGGA..... 564
||||| : : : : : ||| : : : : : ||| : : : : : |||
268 GlyAlaGluTyrTrpArgAspTyrLeuLysLeuSerSerAsnGlyTyrLe 284
565 .AGAATACAGCAGTAGTAGTGTTCAGCAGGATTTACTGTGACAAATGTAG 613
||||| : : : : : ||| : : : : : ||| : : : : : |||
284 uArgLeuThrAsnTrpArgSerAlaProGluLeuAspAsnAspTyrGluA 301
614 TAAAGAAGTCTCAGGGA..... 630
||| : : : : : ||| : : : : : ||| : : : : : |||
301 laArgProAlaAsnGlyTrpAspValargAlaGluGlyTrpLeuProAla 317
631 .....GGCAGCTTTACACCAAAAGATGCACATTTCAATAAAGT 668
||||| : : : : : ||| : : : : : ||| : : : : : |||
318 TrpProHisLeuGlyGlyLysLeuValTyrGluGlnTyrTyrGlyAspG1 334
669 AACAGACTCTATGAAAGAGCAGTGTGAGTTGTCTCCAAATCCGCCAGA 718
||||| : : : : : ||| : : : : : ||| : : : : : |||
334 uValAlaLeuPheAspLys.....AspAspArgGlns 345
719 CGGAGAAGGCTTCTATAATGTTTGCACACATGTTGAT.....TCTATA 762
||| : : : : : ||| : : : : : ||| : : : : : |||
345 erAsnProHisAlaIleThrAlaGlyLeuAsnTyrTrpPheProLeu 361
763 GTTGAATTTCTGACAGACAAAACCAACAAAGAGCTCCAAACCAAGCA 812
||| : : : : : ||| : : : : : ||| : : : : : |||
362 MetThrPheSerAlaGluGlnArgGlnGlyLysGlnGlyGluAsn.... 376
813 AAATCAAAATGCAATCTCCGAGCAGCATCGGAA..... 846
||| : : : : : ||| : : : : : ||| : : : : : |||
377 .AspThrArgPheAlaValAspPheThrTrpGlnProGlySerAlaMetG 393
846 ..... 846
393 lnLysGlnLeuAspProAsnGluValAspAlaArgArgSerLeuAlaGly 409
847 .....GTGATCCGTGATTTCTGAG.....GACTTTAA 872
||| : : : : : ||| : : : : : ||| : : : : : |||
410 SerArgPheAspLeuValAspAlaArgAsnAsnIleValLeuGluTyrAr 426
873 GAAA.....ACCACTCCTATGACA.....A 892
||| : : : : : ||| : : : : : ||| : : : : : |||
426 gLysLysGluLeuValArgLeuThrLeuThrAspProValThrGlyLys 443
893 CACAGCCCAAAATCCACCTTCTCATTTGCTGTCAGATT..... 930
||| : : : : : ||| : : : : : ||| : : : : : |||
443 erGlyGluValLysSerLeuValSerSerLeuGlnThrLysTyrAlaLeu 459
931 ..GGACAAGAATTTGTGTTGTAGTCTTGACAAATCTGGAAGC..... 972
||| : : : : : ||| : : : : : ||| : : : : : |||
460 LysGlyTyrAsnValGluAlaThrAlaLeuGluAlaAlaGlyGlyLysVa 476
973 .ATGGCGACTGGTAACCGGCTC..... 993
||||| : : : : : ||| : : : : : ||| : : : : : |||
476 lValThrThrGlyLysAspIleLeuValThrLeuProAlaTyrArgPheT 493

```

```

993 ..... 993
493 hrSerThrProGluThrAspAsnThrTrpProIleGluValThrAlaGlu 509
994 .....AATCGACTGAATCAAGCAGGACGCTTTTCTCTGCTGCA 1031
||| : : : : : ||| : : : : : ||| : : : : : |||
510 AspValLysGlyAsnPheSerAsnArgGluGlnSerMetValValValG1 526
1032 GACAGTTGAGCTGGGTGCTTGGGTGGGATGGTGCACATTTTACAGAGTCTG 1081
||| : : : : : ||| : : : : : ||| : : : : : |||
526 naLaProThrLeuSerGln.....LysAspSerServ 537
1082 CCATGATACAAAGTGAACATACAGATAACAGTGGCAGTGACAGGGAC 1131
||| : : : : : ||| : : : : : ||| : : : : : |||
537 alSerLeuSerSerGlnThrLeuSerAlaAspSerHisSer..... 550
1132 ACACCTGCCCAAAAGATTACCTGCAGCAGCTTCAGGAGGGACGTCCTCATCTG 1181
||| : : : : : ||| : : : : : ||| : : : : : |||
551 ThrAlaThrLeuThrPheIleAlaHisAspAlaAlaGlyAsnProValI1 567
1182 CACGGGCTTTCGATCGGCATTTACTGTGATTAGGAAGAAATATCCAACTG 1231
||||| : : : : : ||| : : : : : ||| : : : : : |||
567 e...GlyLeu.....ValLeuSerThrArgHis.....G 576
1232 ATGATCTGAAATTTGCTCTGACGGATGGGAGACACACACTATAGT 1281
||| : : : : : ||| : : : : : ||| : : : : : |||
576 luGlyValGlnAspIleThrLeuSerAspTrpLysAspAsnGly...Asp 591
1282 GGTGTGTTTAAACGAGGTCAAAAGTGGTGCATCATCCACACAGTCGC 1331
||| : : : : : ||| : : : : : ||| : : : : : |||
592 GlySerTyrThrGlnIleLeuThrThrGlyAlaMetSerGlyThrLeuTh 608
1332 TTTGGGGGCC.....TCTGCAGCTCAAGAAGTAG 1360
||| : : : : : ||| : : : : : ||| : : : : : |||
608 rLeuMetProGlnLeuAsnGlyValAspAlaAlaLysAlaProAlaVal 625
1361 AGCAGCTGTCCAAATGACAGGAGGTTTACAGACATATGCTTCAGATCAA 1410
||| : : : : : ||| : : : : : ||| : : : : : |||
625 alAsnIleIleSerValSerSer...ArgThrHisSerIleLys 640
1411 GTTCAGAACAAATGGCTCATTTGATGCTTTTGGGGCCCTTTCATCAGGAAA 1460
||| : : : : : ||| : : : : : ||| : : : : : |||
641 IleAspLysAspArgTyrLeu.....SerGlyAs 650
1461 TGGAGCTGTCTCTCAGCGTCCATCCAGCTTGAGAGTAAGGATTAAACC 1510
||| : : : : : ||| : : : : : ||| : : : : : |||
650 nProIleGluValThrValGluLeuArgAspGluAsnAspLysProVal 667
1511 TCCAGAACACCCAGTGGATGAATGGCAGCTGATCGTGGAC..... 1551
||| : : : : : ||| : : : : : ||| : : : : : |||
667 ysGluGlnLysGlnLeuAsnThrAlaValSerIleAspAsnValLys 683
1552 .....AGCACCTGGGAAAGGACACT.....TTGTTTCTTAT 1583
||| : : : : : ||| : : : : : ||| : : : : : |||
684 ProGlyValThrThrAspTrpLysGluThrAlaAspGlyValTyrLysAl 700
1584 CACTCGACAAACGAC.....CCTCCCAAAATCCTTC 1615
||| : : : : : ||| : : : : : ||| : : : : : |||
700 aThrTyrThrAlaTyrThrLysGlySerGlyLeuThrAlaLysLeuLeuM 717
1616 TC.....TGGGATCCAGTGGAGCAAGCAAGTGGCTTTGTAGTGGAC 1659
||| : : : : : ||| : : : : : ||| : : : : : |||
717 etGlnAsnTrpAsn...GluAspLeuHisThrAlaGlyPheIleLeuAsp 732
1660 AAAAACACCAAAATGGCCTACCTCCAAATCCAGGCAATTCCTAAGGTGG 1709
||| : : : : : ||| : : : : : ||| : : : : : |||
738 AlaAsn.....ProGlnSerAlaLysIleAl 741
1710 CACTTGGAAATACAGT.....CTGCAAGCAAGCTCAC 1741
||| : : : : : ||| : : : : : ||| : : : : : |||
741 aThrLeuSerAlaSerAsnAsnGlyValLeuAlaAsnGluAsnAlaAla 758
1742 AACCTTGACCCCTGACTGTGACGTCGCCGTGCTCAATGCTACCCCTGCCT 1791

```

[illegible]

seq_documentation_block:
probable lipoprotein Cj1678 [imported] - Campylobacter jejuni (strain NCTC 11168)
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C:Accession: C81265
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, T.; Whitehead, S.; Barrell, A.G.; Paulsen, O.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barton, G.J.; Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, T.; Whitehead, S.; Barrell, A.G.; Paulsen, O.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barton, G.J.
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barton, G.J.; Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, T.; Whitehead, S.; Barrell, A.G.; Paulsen, O.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barton, G.J.
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: A81250; MUID:20150912
A:Accession: C81265
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-928 <PAR>
A:Cross-references: GB:AL113079; GB:AL111168; NID:g9698971; PIDN:CAB73665.1; RID:g9698971
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj1678

298 GTGAGACAAACTTGAGACCCTACAATAATCGTAGTGTTCGTGCCTCA 347
::: :::::::::::::::::::: :::
50 lleAsnThrLysIleAspTrpPheThrAsnAsnGlyPheIleAsnSerP 66
::: :::::::::::::::::::: :::
348 GNCtAcTcCtCCAgTAATgAtGAACccGTACAcTgAcGCaCATGGCAAcT 397
::: :::::::::::::::::::: :::
66 odglySergLyGlNTrpAsnAsnGlyLeTrp.lIeSerAsnAlaThr 82


```
1480 TCCATCCAGCTTCAGAGTAGGGAATTAAACCTCCAGAACAGCAGTGGAT 1529
||||| : : : : :
3526 SerValAlaValAsp.....Le 3531
1530 GAATGCCAGAGTTCAGTGGACAGACACCGTGGAAAGGACACTTTGTTTC 1579
: : : : : |||
3531 uThrAlaProValIleSerIleAsnThrValSerThrAspAspArg**A 3548
|||
1580 TTATCACCTGGACAACGACCGCTCCCAAAATCCTCTCTGGGATCCAGT 1629
: : : : : |||
3548 snAlaAlaGluGlnGlnPro.....LeuThrLeuAsnGlySerThr 3562
1630 GGACAGAACAGAGTGGCTTTGTAGTGGCAAAAACACAAAATGGCCTA 1679
: : : : : |||
3563 SerAlaGluValGlyGlnThrValThrValThrPheGlyGlyLysThrTy 3579
CTCCAAATCCAGGAGTTCCTAAGCTTGGCACTTGGAAATACAGTCTGC 1729
: : : : : |||
3579 rThrAlaThr.....ValAlaAlaAsnGlyThrTrpAlaLeuAsnValp 3594
1730 AAGCA.....AGCTCACAACCTTTGACCCGTG 1755
|||
3594 roAlaValAspLeuAlaLeuGlyGlnGlyAlaGlnThrIleThrAla 3610
: : : : : |||
1756 ACTGTACGTCGCGTCCGTCCTCAAT..... 1779
: : : : : |||
3611 SerValAsnAspArgAlaGlyAsnProGlyGlnAlaThrHisAlaLeuTh 3627
1780 .....GCTACCTCCCTCCATTACGTACTTCCAAACGACCAAGG 1822
3627 rValAspThrValAlaProThrValThrIleAlaThrValAlaGlyAspA 3644
|||
1823 ACACAGCAAAATTTCCCAAGCCCTCTCGTATGTTTATGCAAAATATTCGCCAA 1872
|||
3644 sp.....IleIleAsnAsnAlaGluGln 3651
1873 GGAGCTCCCAATTCAGGGCCAGTGTACAGCCCTGATTGAATCAGT 1922
||||| : : : : :
3652 LeuAlaGlyGlnThrIleSerGlyThrThrAlaGluVal..... 3665
1923 GAATGAAACACAGTTACCTTGGAACTACTGGATAATGAGCAGGTCGTG 1972
||||| : : : : :
3666 ....GlyGlnThrValThrValThrPhe.....AsnGlyGlnThrTrpS 3679
1973 ATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTTCAAACTTATGAC 2022
||||| : : : : :
3679 erAlaThrValGlySerGly..... 3685
2023 ACGAATGGTAGATACAGTGTAAAGTTCGGGCT.....CTGGGAGGAGT 2066
||| : : : : : |||
3686 ....GlySerTrpSerValPheIleProAlaGlnGlnPheAlaGlyLe 3700
2067 TAACGACGACGACGAGAGTATACCCCGACGAGAGTGGGCACTGTACA 2116
: : : : : |||
3700 uSerAspGlySerTyThrIleSerAlaThrValSerAspGlnAlaGlyA 3717
2117 TACCTGGCTGGATTGAGATGAT...GAATACAAATGGAATCCACCAAGA 2163
||||| : : : : : |||
3717 snProGlySerAlaSerArgGlyValThrLeuAsnGlyAspValProThr 3733
2164 CCTGAATTAATAAG.....GATGATGTTCAACACAGCAAGCAAGTGTG 2204
|||||
3734 ValThrIleAsnThrPheAlaGlyAspAspVal..... 3744
2205 TTTCAGCAGAACATCCTCGGGAGGCTCATTTGTGGCTTCTGTATGCCCAA 2254
: : : : : |||
3745 .ValAsnAlaAlaGluHisGlySerSerLeuValIleSerGlyThrThrT 3761
2255 ATGCTCCCATACCTGATCTCTTCCCACTGGCCAAATCACCACCTGTGAAG 2304
: : : : : |||
3761 hrAlaProVal.....GlyGln.....ThrLeuThr 3769
```

```
2305 GCGGAATTCACGGGGCAGTCTCATTAATCTGACTTGGACACGCTCCTGG 2354
: : : : : |||
3770 LeuThrLeuAsnGlyLysThrTyThrThrValGlnThrGlyGlyse 3786
2355 GGATGATTATGACCATCGAACAGCT.....C 2380
: : : : : |||
3786 rTrpSerTyThrLeuGlySerAlaAspValThrAlaLeuAlaAspGlyA 3803
2381 ACAAGTATATCATTCGAATAAGTACAAAGTATTCTTCATCTCAGAGCAAAG 2430
: : : : : |||
3803 snAlaTrpValIleAsnAlaSerValSer.AsnAlaIleGlyAsnThrGI 3819
2431 TTCAATGAAT 2440
: : : : : |||
3819 ySerSerAsn 3822
seq_name: pir2:B71704
seq_documentation_block:
cell surface antigen (sca3) RP451 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: B71704
R:Anderson, S.G.E.; Zomorodipour, A.; Anderson, J.O.; Sicheritz-Ponten, T.; Alsmark
Naturae, 1996, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499
A:Accession: B71704
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2340 <AND>
A:Cross-references: GB:AJ235271; GB:AJ235269; NID:9368717; PIDN:CAA14908.1; PID:g386
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: sca3; RP451
alignment_scores:
Quality: 143.50 Length: 809
Ratio: 0.395 Gaps: 38
Percent Similarity: 44.870 Percent Identity: 20.025
alignment_block:
US-09-049-696-20 x B71704 ..
Align seg 1/1 to: B71704 from: 1 to: 2340
513 GGAGTATTTGACGAGTACAATATGATGAGAAATTCCTACTTATCCAAATGG 563
||| : : : : : |||
840 GlyAspIleAspPheAsnAsnLysAlaGlyLysPheIleLeuGlyAspGI 856
563 AAGAATA.....CAAGCAGTAAGATGTTTCAGCAGGATTTACTTGGTACAA 607
| : : : : |||
856 yAlaMetIleAspGlySerValLeuCysAsnGlyGlyValAlaGlyThr. 872
: : : : : |||
603 ATGTAGTAAGAAGTGTCTCAGGGAGGAGCTGTTACACCAAAAGATGCACA 657
872 ..... 872
658 TTCAATAAAGTAACAGGACTCTATGAAAAGGATGTGAGTTTCTTCTCCA 707
: : : : : |||
873 .....LeuAspPheIleGlyAs 878
708 ATCCCGCCAGACGAGGAGGCTTCTATAATGTTTTCGCAACATGTTGATT 757
: : : : : |||
878 pGlyAsnValThrGlnAsnIleGlyAlaAspAsnAlaAsnSerIleSert 895
758 CTATAGTTGAATTTCTGTACAGAACAAACCAACAAAGAGCTCCAAAC 807
: : : : : |||
895 hrIle.....AsnIleGlnGlyAspAsn 902
803 AAGCAAAATCAAAAATGCAATCTCCAGACCATGGGAAGTGTATCCTGTA 857
: : : : : |||
```

903 ThrLys.....AsnValThrIleAlaAsnAspIlePheValas 915
 858 TTCGAGGACTTTAAGAAACCACTCTCTATCACAACACAGCCCAAAATC 907
 I::: III::: :
 915 pAsnIleHisPheThrAsnGly..... 922
 908 CCACCTCTCTCATTCGTCGAGATTGGACAAGAATTGTGTTTAGTCTCTT 957
 :::: :::: :
 923GlyIleLeuGlnLeuGlyGlyAsnLeuThrThrHisAsnIle 936
 958 GACAATCTGGAAGCATGGCGACT.....GGTAACCCGCCCA 995
 III::: :::: :
 937 AspPheGlyAlaAsnGlyGlyThrLeuGluPheAsnGlyAsnAsnThrTy 953
 996 TCGACTGAATCAAGCA.....GGCCAGCTTTTCTGCTGCAGACAG 1036
 :::: :::: :
 953 rAsnLeuAsnAlaIleIleValasnGlyGlnAsnGlyIleLeuAsnAlap 970
 1037 TT.....GAGCTGGGGTCTGGTGGTGGATG 1062
 :::: :::: :
 970 heThrAsnLeuLysAlaSerAspThrIleGlyThr...ValLysIle 985
 1063 GTGACATTGTGACAGTCTGCCATGTACAAAGTGAACATACAGATAAA 1112
 :::: :::: :
 986 IleAsnIleGlyGlnIleGlyThrProGlnAsnPheThrIleGlnValas 1002
 1113 CAGTGGCAGTGACAGGACACACATCGCCAAAAGATTACTGCAGCAGCTT 1162
 I::: :::: :
 1002 nAsn.....LysAsnLeuThrLeuValSers 1011
 1163 CAGGAGGAGCTCATCTGCACGGG..... 1188
 II::: :::: :
 1011 erValAsnSerSerIleAsnPheGlyAspAlaAsnSerGlnLeuIleLeu 1027
 1189CTTCGATCGGCATTACTGTGATTAGGAAGAAATATCCAAC 1229
 :::: :::: :
 1028 SerAlaProValAspGlnThrIleLysPheIleAsnAsnLeuAsnGluTh 1044
 1230 TGATGATCTGAAATTGCTGCTCAGCATGCGGAGGAGCAACACTATAA 1279
 :::: :::: :
 1044 rGlyGlyGlyIleIleThrLeuAspSerAsnGlyAsnAsnLeuThrIles 1061
 1280 GTGGTGTCTTACGAGGTCAACAAAGGTGGCCATCATCCACACAGTC 1329
 III::: :::: :
 1061 erGly.....AsnAsnGly.....Ile 1066
 1330 GCTTTGGGCCCTCTGCAGCTCAAGAACTAGAGAGCTGTCTCAAATGAC 1379
 :::: :::: :
 1067 LysLeuGly...SerLysGlyAsnGluLeuSerSerLeuAsn.....II 1080
 1380 AGGAGTTTACAGACATATGCTTCAGAT...CAAGTTCAGACAATGGCC 1426
 :::: :::: :
 1080 eLysGlyLysValThrValThrAsnAspLeuAspIleGlnAsn..... 1094
 1427 TCATTGATGCTTTGGGCCCTTTCATCAGGAATGGAGCTGTCTCTCAG 1476
 :::: :::: :
 1095IleHisGlnLeuAsnIleAsnAsnGlyAlaLeu..... 1105
 1477 CGCTCCATCCAGCTTGAGACTAAGGATTAACTCCCAAGACAGCAGTC 1526
 :::: :::: :
 1106PheAspGlnSerLeuThrSerAlaLysIleLysAs 1118
 1527 GATCAAT...GGCAGCATGATCGTCAGACAGCAGCTGGGAAAGACACTT 1573
 :::: :::: :
 1118 nIleAsnIleGlyThrValAlaGlyGlyAlaThrThrThrLeuAspAlaI 1135
 1574 TGTTTCTTATCATCTGGACAACGACGCTCCCAATCTCTCTGGGAT 1623
 :::: :::: :
 1135 le..... 1135
 1624 CCCAGTGCAGAGCAGAGTGGCTTTGTAGTCGACAAAACACCAAA... 1671
 :::: :::: :
 1136AsnAspAsnPhe.....AspLeuAsnThrSerGI 1145

1672 .ATGGCCTACCTCCAAATCCCAAGGATTCCTAAGGTTGGCAGCTTGAAT 1720
 :::: :::: :
 1145 yMetValPheLysHisGlnAspSerIleLeuGluLeu.....LysA 1159
 1721 ACAGCTGCAAGCAAGCTCACAACCTTGACCCCTGACTGTCACTGCCGT 1770
 :::: :::: :
 1159 snSerSerAsnThrAsnAspHisThrIleThrLeuThrSerAlaLeuAsp 1175
 1771 GCGTCAATGCTACCTCGCTCCCAATACAGTGACTTCCAAACGACAA 1820
 :::: :::: :
 1176 ProGlyAsnAsnGlnPheGlyIleIleLysLeuIleThrAspThrAsnLy 1192
 1821 G.....GACACAGCA 1831
 1192 sLeuThrIleAspAsnAsnGlyAsnValAlaTyrThrLeuGlyThrAlaA 1209
 1832 AATTCCCCAGCCCTCGTAGTTTATGCAATATTCGCCAAGGAGCCTCC 1881
 :::: :::: :
 1209 snHisMetLeuLysGlnLeuThrPheAlaSerIleAspAsnGlyAla... 1224
 1882 CCAATTCTCAGGCCAGGTGTACAGCCCTGATTGAATCACTGAATGAAA 1931
 :::: :::: :
 1225IleAlaLeuLysValGlyIleAsnValGI 1234
 1932 AACAGTTACCTTGGAACTACTGGAT..... 1956
 :::: :::: :
 1234: uAsnValThrLeuAsnIleLysAspIleGluLeuAsnGluValAsnAlaA 1251
 1957AAT 1959
 1253: snValLeuPheAsnLysAsnThrThrTyrThrAlaThrGlyAsnIleAsn 1267
 1960 GGA.....GCAGGTGCTGATGCTACTACTAAGATGA 1988
 :::: :::: :
 1268 GlyHisValAspPheGlnGlyAsnAlaGlyValIleAsnLeuAsnAspAs 1284
 1889 C...GGTGTCTACTCAAGGTATTTCACACTTATGACACGAATGGTAGAT 2035
 :::: :::: :
 1284 pIleGluIleAspGlySerValThrSerThrGlyAsnValAsnGly.... 1299
 2036 ACAGTGTAAAGTGGGCTCTGGAGGAGTTAACGCCAGCAGCAGGAGA 2085
 :::: :::: :
 1300...ThrLeuAsnPheAsnGlySerGlyLysValThrGlyLeuIleAsnAsn 1315
 2085 GTGATACCCCGACAGAGTGGAGCA.....CTGTACATACCTGG 2123
 :::: :::: :
 1315 IleValMetLeuGlnAlaGlyAlaGlyAspValSerLeuSerAlaSerGI 1332
 2124 CTGGATTGAGAATGATGAATACAAATGAATGCCACCAAGA..... 2163
 :::: :::: :
 1332 yAsnTyrSerIleThrGluIleGlnGlyAsnGlyAsnAsnLeuThrP 1349
 2164CCTGAAATTAATAGGATGATGTTCAA 2190
 1349 heAlaAlaAsnSerHisLeuThrThrAspIleAsnLysThrGlyGlyGln 1365
 2191 CACAAGCAAGTGTGTTCAGCAGACAATCCTCGGGAGGCTCATTTGTCG 2240
 :::: :::: :
 1365 AspLeuAsnLeuValPhe.....IleAsnGlyGlySerValSerGI 1379
 2241 TTCGATGTCCTCCAAATGCTCCATACCTGATCTCTCCACCTGGCCAAA 2290
 :::: :::: :
 1379 ySerIleGlyAlaAsnAlaAlaValGlyAspIleIleIleAsnAlaGlyS 1396
 2291 TCACGAGCTTGAAGCGGAAATTCACGGGGCAGTCTCATT..... 2331
 :::: :::: :
 1396 erValAsnPheSerAsnThrLeuLysSerGlyAsnIleValIleSerAsp 1412
 2332AATGACTTGGACA...GCTCCTGG 2354
 :::: :::: :
 1413 GlyAlaThrMetGlnValAsnAsnValThrAlaThrAspIleSerGI 1429

186 GGACATGGTGACCCAGGCATCTCTGTATCTGTTGAAGCTACAGGAAAGC 23

304 uGlyArgGlyValGlyIleLeuAsnArgSerGlyLysIleGluGlyS 321
1028 TGCAGACAGTGTACCTGGGTCTCGGTT..... 1056
321 erileLysValGluAspGlyAlaThrValAlaThrSerAsnArgAla 337
1057GGGATGGTGACA..... 1068
338 IleAlaAsnSerGlySerGlySerIleThrGlyIleThrValSerG1 354
1068 1068
354 yLysAsnThrLysLeuGluGlyAsnIleIleAsnThrGlyAsnAlaSerI 371
1069TTTCACAGTGTCCCATGTACAAAGTGAACCTC 1101
371 leGlySerAspIleLysIleGlyAlaLysValGluGlyGlyLeu 387
1102ATACAGATAAACAGTGG 1118
388 ValAsnGlnGlyAsnGlySerIleSerGlySerValGlnValSerGlyG1 404
1119 CAGTCACAGGACACACTCGCCCAAGATTACCTGCACGACCTTCAGGAG 1168
404 ySerSerIleAspSerIleThrAsnGluGlyAsnGlyAlaIleSerGlyS 421
1169 GGACCTCCATCTGC.....AGCGGCTTCGATCGGCATTTACTGTGATT 1212
421 erileThrValIryLysAspSerLysLeuAspSerIleThrAsnThrSer 437
1213 AGGAAGAAATATCCAACTGGATGTGAATTTGCTGTGACGGATGG 1262
438 ThrSerSerThrGlyIleSerGlySer.....IleThrAsnAs 450
1263 GGAACACACACTATAAGTGGTGTTTAACAGAGCTCAACAAAGTGTG 1312
450 nSerAspAsnLysLeu.....GluIleSerAsnSerGlyA 462
1313 CCATCATCCACACAGCTCGCTTTGGGGCCCTCTGCAGCTCAAGAAGTAGAG 1362
462 snIle..... 463
1363 GAGCTGTCCAAATACACAGGAGTTA.....CAGACATATGCTTCAGA 1406
464GlyGlyLysIleGluSerThrGlySerAlaAs 474
1407 TCAAGTT.....CAGAACAAATGGCTCATTTGATGCTTTTGGGGCC 1447
474 pMetValIleSerAsnSerAsnGlyGlyThrIleSerGly...GlyIles 490
1448 TTTTCATCAGAAATGAGCTGTCTCTCAGCGCTCCATCCAGCTTACAGAT 1497
490 erSerSerGlySerGlySerThrIleSerAsn.....Ser 502
1498 AAGGGATTAACTCCAGAACACAGCAGTGGATGATGGACAGTATGATGT 1547
503 GlnGlySerThrIleAsnAsnGlyIleThrValSerGlySerAlaGlnVa 519
1548 GGAC.....AGCACCGTGGGAAGGACACTTTGTTCATTATCA 1585
519 lGluIleSerAsnGlnGlySerValGlyLysAsp..... 530
1586 CTGGACACAGCGACCTCCCAATCTTCTCTGGGATCCAGTGACAG 1635
530 530
1636 AAGCAAGGTGGCTTTGTAGTGACAAAAACACCAAAATGGCTTACCTCCA 1685
531GluAsnGlyAsnThrValThrAsnAsn..... 539
1686 AATCCCAAGCATTTGCTAAGTTGGC.....ACTTGGAAATACAGTCTOC 1729
540GlySerGlySerValGlyIleLysAspTrpLeuValSerThra 554

1730 AAGCAAGCTCACAAACCTTGACCCCTGACTGTCCAG.....TCCCGTCCG 1773
554 sPlYAsnThrGlyLysLeuAsnThrValIleGlySerArgAla 570
1774 TCATGTACCTGCTCCCTCAATTCAGTGTCTCCAAAGCAACAAGGA 1823
571 PheAsnValLysValGluAsnIleThrVal...AspGlnSerAsnValAs 586
1824 CACCAAGAAATTTCCCAAGCCCTCTGTAGTTTATGCAAAATATTCGCCAAG 1873
586 pleuGluGluLeu..... 590
1874 GAGCTTCCCAATTTCTCAGGCCAGTGTACAGCCCTGATTTGAATCAGTG 1923
591AsnAspIleAsnAsnIleIleSerGlyVal 600
1924 AATGGAANAACAGTTACCTTGGAACTACTGGATAATGGAGCAGGTGCTGA 1973
601 AsnGlnAsnAsnIleGly...AsnIleGlyThrAsnGlySerGlyGluI1 616
1974 TGCTACTAAAGGATGACGGTCTCTACTCAAGGTATTTTCACAACCTTATGACA 2023
616 eSerLeuSerPheAspProIleThrGlyLysLeuThrThrAspPheAsnL 633
2024 CGAATGGTAGATACAGT...GTAAAGTCCGGGCTCTGGGAGGAGTTAAC 2070
633 euAsnAlaSerIleSerGlyAlaThrPheArgSerLeuIleSerThrThr 649
2071 GCAGCCAGCAGGAGATGATACCCAGCAGAGTGGAGCAGCTGTACATACC 2120
650 SerArgArgSerThrPheIle..... 656
2121 TGCTGGATTGAGATGATGAATAATACAAATGGAATCCACCAAGACCTGAA 2170
657AspAsnValMetGlyAsn..... 662
2171 TTAATAGGATGATGTTCACACACAGCAAGTGTGTTTCAGCAGAACATCC 2220
663SerMetGlnSerPheAlaLeuAlaSerSerSer 673
2224 TCGGGAGGCTCATTTGTGGCTTCTGATGTCCCAAT.....GCTCCCAT 2264
674 LysSerGlnSerIleAlaMetSerGluLysGlyAsnLeuThrAlaAspAl 690
2265 ACCTGATCTCTCCACCTGCCCAATACCCAGCTGAAGCGGAATTC 2314
690 aSerAspTyr.....IleLysSerAspLeuA 699
2315 ACGGGGAGCTCTCATTAATCTGACTTGGACAGCTCTCTGGGATGATTAT 2364
699 snAsnGlySer.....Tyr 703
2365 GACCATGGACAGCTCACAATATATCATTCGATATAGTACAGTATTTCT 2414
704 GlySerAsnLysGluHisSerLeuPheIleLeuProThrThrSerSerG1 720
2415 TGATCTCAGACACAAAGTTCATGAATCTCTCAAGTCAATACTACTGCTC 2464
720 nasnValGluLeuSerLeuAsnGluSerLysGlyHisThrLysGlyT 737
2465 TCATC.....CCAAAGGAAGCAACTCTGAGGAAGTCTTT 2499
737 hrIleIleGlyTyrSerThrLeuLysAspSerGlyIleTyrGlyValTyr 753
2500TTGTTTAAACACAGAAA 2516
754 AlaGlyTyrGluAspThrLysMetGlySerThrTyrPheAspIleAsnAs 770
2517 CATTAATTTT.....G 2527
770 nArgThrTyrTyrAlaGlyLeuLysTyrPheAsnThrLeuPheThrThrG 787

2528 AAAATGCGACAGATCTTTTCATTCGTTATTCAGGCTGTTGAT 2568
|||||
787 luyysGlyGlnGluValThrIleLysAlaGlnGlyLysAlaAlaLeu 803
|||
2569 AAGGTGCATCTGAATCAGAAATATCAAC 2598
|||
804 LysAsnAspLeuThrGluLysIleGlyAsn 813

seq_name: p1r2:A43855

seq_documentation_block:
high-molecular-weight surface-exposed protein HMW1 - Haemophilus influenzae

C:Species: Haemophilus influenzae
C:Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999

C:Accession: A43855

R:Barenkamp, S.J.; Leininger, E.

Infect. Immun. 60, 1302-1313, 1992

A:Title: Cloning, expression, and DNA sequence analysis of genes encoding nontypeable Haemophilus pertussis.

A:Reference number: A43855; MUID:92192797

A:Accession: A43855

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1536 <BAR>

A:Cross-references: GB:U08876; GB:M4616; NID:g475770; PIDN:AAA20527.1; PID:g475771

A:Note: sequence inconsistent with the nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBI:89235, NCBI:89239)

alignment_scores:

Quality: 138.50 Length: 936

Ratio: 0.318 Gaps: 49

Percent Similarity: 46.581 Percent Identity: 20.406

alignment_block:

us-09-049-696-20 x A43855 ..

Align seg 1/1 to: A43855 from: 1 to: 1536

217 TTTCAGCTTACAGGAACCGATTTTATTTTCAAAATGT 254
|||||
631 PheGluGlyThrLeuAsnIleSerGlyLysValAsnII 643
255 TGCATTTTCATCTCAACATGGAGACAAAGCGTGACTATGTGAGAC 304
|||||
643 eSerMetValLeuProLysAsnGluSerGlyTyrAspLysP 657
305 CAAACCTTGAGACTACAAAATGCTGATGTTCTG...GTGCTGAGTCT 351
|||||
657 hElyGlyArgThrTyrTrpAsnLeuThrSerLeuAsnValSerGluSer 673
352ACTCTCCAGGTAATGAT 369
674 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspSerAlaGlyTh 690
370GACCCCTACACTGACGAGA 388
690 rLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThr 707
389 TGGGCAACTCTGGAGAGAAGGCTGAAAGGATCCACCTCACTCCT 432
|||||
707 hrPheAsn.ValGluArgAsnAlaArgValAsnPheAsnIleLysAlaPr 723
433 GATTTCATTCAGGAAAGAAAGTTAGCTGAATATGGACCACAAAGGTAGGC 482
|||||
723 oIleGlyIleAsnLysTyrSerSerLeuAsnTyrAlaSerPheAsnGlyA 740
483 ATTTGTCATGAGTGGCTCATCTACATGGGGAGTATTGGACGAGTACA 532
|||||
740 snIleSerValSerGlyGlyGlySerValAspPheThr 752
533 ATATGATGAGNAATCTACTTATCCAAATGGAGAATACAAGCA 576
|||||
|||||

753 LeuLeuAlaSerSerSerAsn.ValGlnThrProGlyValValIleAsnS 769
|||||
577 GTAAGATGTTACAGCATATTACTGTCACAAATGTAGTAAAGAGTGTCA 626
|||||
769 erLysTyrPheAsn.ValSerThrGlySerSerLeuArgPheLysThrSe 785
627 GGCAGGCAGCTGTTACACCAAAAGATGCACATTCAAT. 663
|||||
785 rGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeuAsnA 802
664AAAGTAACAGAGCTCTATGAAAAAGCATGTGAGTTGTT 702
|||||
802 laThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspGlyMet 818
703 CTCCAATCCGCCAGCAGCGAGAGGCTTCTATAATGTTT 741
|||||
819 IleGlyLysGlyIleValAlaLysAsnIleThrPheGluGlyGlyAs 835
742GCACAACATGTTGATTCTATAGTTG 766
835 nIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsnValT 852
767 AATTCTGTACAGAACAAAAC 786
852 hrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPheAspAsn 868
787 CACAACAAAGAGCTCCAAACAGCAAAATCAAAATGCAA 827
869 HisGlnLysProLeuThrIleLysLysAspValIleAsnSerGlyAs 885
823 TCCTCCGAAGCACATGGGAAGTGATCC 853
885 nLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuThrValG 902
854 GTGATTCTGAGGACTTTTAAGAAAACCACTCTCTATGACA 891
902 luSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsnValGly 918
892ACACAGCCACCAATCCACCTTCTCATTTGTCGAGATTGG 932
919 GlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLysGlyI 935
933 ACAAGAATTTGTGTTTAGTCTCTGACAAATCTGGAAGCATGCGACTG 982
935 YAlaArgPheLysAspIleAspAsnSerLysAsnLeuSerIleThrThrA 952
983 GTAACCGCCTCAATCGACTCAATCAA 1008
952 snSerSerSerThrTyrArgThrIleIleSerGlyAsnIleThrAsnLys 968
1009 GCAGGCCAGCTTTTCTGTCGACAGACAGTTGAGCTGGGTCCTGGGTGG 1058
969 AsnGlyAspLeuAsnIleThrAsnGluGlySer 979
1059 GATGGTGCACATTTGACAGTCTGCCCATGTACAAAGTGAACCTCATACAG. 1107
980AspThrGluMetGlnIleGlyAspValSerGlnL 992
1108ATAACAGTGCAGGTGACAGGACACACCTCGCCAAAGA 1146
992 ysGluGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrLysGln 1008
1147TTACCTGCAGCAGCTTCAGGAGGACGCTCCATCTGCAGCGGCT 1190
1009 IleThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAlaTh 1025
1191 TCGATCGGCATTTACTGTGATTAGGAGAAA 1221
1025 rAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnA 1042
1222TATCCAACCTGATGA 1236
1042 spLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGly 1058

```
1237 TCTGAATGTGCTGACGAGTGGGAGACAAACACTATAAGTGG... 1284
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1059 SerAsp.....LeuThrIleGlyAsnThrAsnSerAlaAspGlyTh 1072
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1285 .....TGCTTTACGAGGTCACAAAGT.....GGTG 1312
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1072 rAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1313 CCATCATCCACACAGTCCCTTTGGGCGCCTCTGCAGCTCAAGAACTAGAG 1362
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1089 laAspGlyHisLysValThrLeuHis..... 1097
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1363 GAGCTGTCCAAATG...ACAGGAGGTTTACAGACATATGCTTCAGATCA 1409
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1098 .....SerLysValGluThrSerGlySerAsnAsnAsnThrGluAspSe 1112
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1410 AGTTTCAGAACAAATGCCTTCATTGATGCTTTTGGGCGCCTTTTCATCAG 1459
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1112 rSerAspAsnAsnAla..... 1117
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1460 ATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGTAAGCGGATTACCC 1509
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1118 .....GlyLeuThrIleAspAlaLysAsnValThr 1127
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1510 CTCCAGAACAGCAGTGGATGAATGGCACAGTGCCTGGAGACACCGGT 1559
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1128 ValAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1560 GGGAAAGGACACTTTGTTCTTATCATCCTCGGACACAGCAGCCTCCCAAA 1609
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1144 rGlyGlu.....IleThrThrLysThr..... 1151
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1610 TCCTTCTCTGGATCCAGTGGACGACGAGGCTGCTTGTAGTGGAC 1659
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1152 .....GlyThrThrIleAsn 1156
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1660 AAAAACACCAAATGGCTTACCTCCAAATCCAGGCATTTGCTAAGTTGG 1709
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1157 AlaThrThrGly.....AsnValGluIleThrAlaGlnThrG1 1169
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1710 CACTTGGAAATACAGTCTCAAGCAGCTCACAAACCTTGACCCGACTG 1759
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1169 ySerIleLeuGlyGlyIleGluSerSerGlySerValThrLeuThrA 1186
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1760 TCAGTCCCGTGG.....TCCATGCTACCTGCCCTCCCAATTACA 1800
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1186 laThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1801 GTGACTTCCAAACGACGAGGACACGAC..... 1830
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleL 1219
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1830 ..... 1830
1219 sGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyLT 1236
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1831 .....AAATTCGCCAGCCCTCTGCTAGTT 1854
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1236 hrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThrThr 1252
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1855 TATGCCAAATATTCG.....CAAGGAGCCTCCCAATTCCTCAG 1892
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1253 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 1269
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1893 GGCAGCTGCACGCCCTGATTGAATCAGTGAATGGAACAAAGTTACCT 1942
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1269 rAlaThrGlyThr...IleGlyGlyThrIleSerGlyAsnThrValAsnV 1285
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1943 TGGNA.....CTACTGGATAATGAGCAGCGTCTGAT 1974
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1285 alThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 1301
```

alignment_scores:
Quality: 138.00 Length: 602
Ratio: 0.531 Gaps: 33

seq_name: pir2:T21810

seq_documentation_block:

hypothetical protein F35G12.4b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T21810

R:Chui, C.

submitted to the EMBL Data Library, October 1994

A:Reference number: Z19473

A:Accession: T21810

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-683 <WIL>

A:Cross-references: EMBL:Z46242; PIDN:CAA86338.1; GSPDB:GN00021; CESP:F35G12.4b

A:Experimental source: clone F35G12

C:Genetics:

A:Gene: CESP:F35G12.4b

A:Map position: 3

A:Introns: 62/3; 105/1; 135/3; 200/3; 269/3; 332/2; 390/1; 425/3; 458/3; 490/1; 542/2

Percent Similarity: 43.189 Percent Identity: 20.598

alignment_block:

US-09-049-696-20 x T21810 ..

Align seg 1/1 to: T21810 from: 1 to: 683

```
1270 AACACTATAGTGGGTGCTTACGAGGTCAACAAGTGGTGCCATCAT 1319
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
172 AsnAsnLeuIleGlyCys.....LysAspSer.....I 181
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1320 CCACAGTCGCTTTGGGCGCTTCAGCTCAAGAACA..... 1359
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
181 eTyrSerLeuAlaThrProAsnLeuSerLeuValLeuAlaGlyT 198
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1360 .....GAGGAGCTGTCC 1371
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
198 hrGluLysCysIleArgLeuPheAspProArgThrAsnGluLysIleMet 214
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1372 AAATGACAGGAGTTTACAGACATATGCTTCAGATCAAGTTTCAGAACAA 1421
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
215 LysLeuArgGlyHisThrAspAsnValArgAlaLeuValValAsnAspAs 231
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1422 TGGCCTCATTGATGCTTTGGGCGCTTCATCAGGAATGGAGTGTC. 1470
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
231 pGlyThr.....ArgAlaLeuSerAlaGlySerAspAlaThrI 244
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1471 .....TCTCAGCGCTCCATCCAGCTTGAGAGTAAG 1500
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
244 leArgLeuTrpAspIleGlyGlnGlnArgCysIle.....AlaThr 257
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1501 GGATTAACTCCAGACACCGCAGTGGATGATGGCACAGTATCGTGA 1550
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
258 CysIleAlaHisGluGluGlyValTrp.....ThrLeuGlnValAs 271
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1551 CAGC.....ACCTGGGAAGGACACTTTCTTCTCTTA 1582
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
271 pSerSerPheThrValTyrSerAlaGlyLysAspLysMetValVal 288
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1583 TCACC.....TGGACAACGACCGCTCCCAATCTCTCTGGGATCCC 1626
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
288 ysthrProLeuTyrAspPheThrLysSerGlnLeuLeuPheLysGlu 304
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1627 AGTGGACAGNAGCAGGTGCTTGTAGTGACAAAAACACCAAAATGCC 1676
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
305 AlaProValLysLys.....LeuLeuLeuLeu 313
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1677 CTACCTCCAAATCCAGGCAATGCTAAGGTTGGCACTTGGAAA..... 1719
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
313 rGluLysAspAsnProValSerLeuTrpValGlyThrTrpLysSerAspI 330
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1720 .....TACAGTCTGCAAGCAGCTCAACAACCTTGACCCCTGACTGTC 1761
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
330 leLysArgTrpSerIleArgProSerAlaGln.....LeuSerIle 343
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1762 ACCTCCCGCTGCTCAATGCTACCTCCCTCCCAATTACAGTACTCCAA 1811
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
344 GlyGlyAspGluAspGlyProSer.....ThrSerAs 354
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1812 AACGAACAGGACACGACCAATATCCCGCCCTCTGGTAGTTTATGCAA 1861
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
354 nAlaAsnHisSerValSerAlaSerSerProProValThrPheLysT 371
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1862 ATATTCCCAAGGAGCGCTCCCAATTTCTCAGGGCCAGTGTCAGAGCCCTG 1911
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
371 yIleArg..... 373
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1912 ATTGAATCAGTGAATGAAACAGTTACTCTGGAACTACTGATATAGG 1961
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
374 ValLysAspGlnLysGlyGlnGlnSerThrProGluLeuValIle.... 388
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1962 AGCAGTGTCTGATGCTACTAAGGATGACGCTGTCTACTCA.....AGGT 2005
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

```
389 ProGlyAlaProAlaIleLysLysHisAlaMetLeuSerAspLysArgH 405
2006 ATTTCAAACTTATGACACAAATGGTAGATACAGTGTAAAGTCCGGCT 2055
|||  |||  |||  |||  |||  |||  |||  |||
405 isValLeuThrArgAspSerAspGlyAsnVal.....Ala 416
2056 CTGGGAGGAGTTAACGCAGCCAGACGG..... 2082
|||  |||  |||  |||  |||  |||  |||  |||
417 LeuTyrAspValLeuAlaAlaArgLysIleLysAspTyrGlyLysArgII 433
2083 .....AGAGTACATACCCAGCAGAGTGGAGCAGCTGTACATACCTGGCT 2125
|||  |||  |||  |||  |||  |||  |||  |||
433 ePheGluGluValValAspGluAsnSerArgGlnValTyrIleProSerT 450
|||  |||  |||  |||  |||  |||  |||  |||
2126 GG.....ATTGAGAATGATGAA 2142
|||  |||  |||  |||  |||  |||  |||  |||
450 rpPheValValAspSerLysSerGlyMetLeuGlnIleThrLeuAspGlu 466
2143 ATA.....CAATGGAATCCACNAGACCTGAA...ATTATATA 2177
|||  |||  |||  |||  |||  |||  |||  |||
467 LeuAspAlaLeuSerSerTrpLeuSerSerLysAspAlaGlyPheAspAs 483
2178 GGATGATGTTTCAACACAAAGCAAGTGTGTTTCAGCAGAACATCTCGGGAG 2227
|||  |||  |||  |||  |||  |||  |||  |||
483 pAsnAspArgGluThrLysGlnIleIle.....ValAsnTyrGlyG 497
2228 GCTCATTTGTGGCTTCTGATGTCCCAATGCTCCCATACCTGATCTCTTC 2277
|||  |||  |||  |||  |||  |||  |||  |||
497 LysMetMetLeuArgSer.....LeuPheGluArgTrp 507
2278 CCACCTGGCCAAATCACCGACCTGAGGGCGGAATTCACGGGGCAGCTCT 2327
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
508 ProProCysLysMetThrAsnValAspAla..... 517
2328 CATTAATCTGACTTGCACACGCTCTCTGGGATGATATATGACCATGGAAC 2377
|||  |||  |||  |||  |||  |||  |||  |||
518 .....AlaAspAlaAspAspValGlnLysAlaThr. 527
2378 CTCACAAGTATATCATTCGAATAAGTACAAGTATCTTGTATCTCAGAGAC 2427
|||  |||  |||  |||  |||  |||  |||  |||
528 .....LeuAsnPheIleSerLeuProGlu 535
2428 AGCTTCAATGAATCTCTCAAGTGAATACTACTGCTCTCATCCCAAGGA 2477
|||  |||  |||  |||  |||  |||  |||  |||
535 His.....ThrProIleIleCysGI 543
2478 AGCCAACTCTGAGGAAGTCTTT.....TTGTTTAAACACGAGAAACATTA 2521
|||  |||  |||  |||  |||  |||  |||  |||
543 uGlyAsnGlyArgProLeuTyrArgLeuLeuValGlyAspAlaGlyLysG 560
2522 CTTTTTCAA...AATGGCACAGATCTTTTTCATTGCTATTTCAGGCTGTGAT 2568
|||||  |||  |||  |||  |||  |||  |||  |||
560 LysPheGluAlaAsnGluLeuAlaGlnIleAlaProMetTrpValIleAsp 576
2569 AAGGTGCTGTAATCAGAAATATCCAACTGACAGGAGTATCTTTGTT 2618
|||  |||  |||  |||  |||  |||  |||  |||
577 AlaIleGlu...ArgAsnGlnLeuProLysPheAsnLysMetProPheTy 592
2619 TATT.....CTCCACAGACTCCCGCAGACACCTAGTCTGTATGATAA 2662
|||  |||  |||  |||  |||  |||  |||  |||
592 rLeuLeuProHisProSerThrAsnProLysGlnProLysLysAspArgL 609
2663 CGTCTGCT..... 2670
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
609 euSerAlaThrGluMetLeuGlnValLysLysValMetGluHisValTyr 625
2571 .....CCTTGTCTTAATAT 2684
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
626 GluLysIleLeuSerThrAsnAspValGlySerIleProLeuAsnGlnII 642
2685 TCATATACACAGCAGCCACTT.....CCTGGCA 2710
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
642 eHisThrLysMetGluMetTyrCysAsnAspGlnArgLeuGluProAspW 659
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

676 LeuLeu 677

[illegible]

alignment_scores:			
Quality:	136.50	Length:	758
Ratio:	0.392	Gaps:	40
Percent Similarity:	45.910	Percent Identity:	21.108
alignment_block:			
US-09-049-696-20 x A85547 ..			
Align seg 1/1 to: A85547 from: 1 to: 1461			
892	ACACAGCCACCAATCCCACTTCTCA.....TT	920	
197	ThrAspThrThrProThrPheSerGlySerGlyGlnAlaGlyAla	213	
921	GCTGCAGATT.....GGCAAGAATTTGTGTTTAGTCCTTG	958	
213	rileGlnIleLysAspSerAsnGlyAsnThrIleAlaSerThrGlnVal	230	
959	ACAAATCTCTGGA.....AGCATGGCGACTGGTAACCGCCTCAAT	996	
230	spAsnAsnGlyHisrpsrValSerLeuProThr.....	241	
997	CGACTGAATCAAGCAGCGCAGCTTTTCCGTGTCGACAGCAGTGTG	1046	

1738TCACAAACCTTGACCCCTGACGTGTCACGT 1765
537 lThrArgThrLeuAspArgGlyAlaAsnThrIlePheValThrValThrA 554
1766 CCCGTGGCTCAATGCT.....ACCTG..... 1788
554 sAlaAlaGlyAsnThrGlyAlaAlaSerArgAlaIleThrLeuValGly 570
1789CCTCAANTACAGTACTCCAAACGAAACGACGACCCACAA 1832
571 ValSerProLeuIleThrIleAsnThrValSerGlyAspAsp..... 584
1833 ATTCGCCGCCCTCTGTAGTTATGCAATATTCGCAAGGAGCCCTCCC 1882
585IleIleSerGlyAlaGluLysGlyAlaProL 595
1883 CAATTCCTCAGGCGCAGTGTACAGCCCTGATTTGAATCAGTGAATGGA 1932
595 euThrLeuThrGlySerThrGln.....Gln**GluThrGlyGln 608
1933 ACAGTTACTTGGAACTACTGGATAATGGACAGTGTCTGATGCTACTAA 1982
609 thrValThrValThrLeu.....AlaGlyGlnSerPheThr.. 620
1983 GGATCAGCGTGTCTACTCAAGGTATTTCCAACTTATGACACGAAATGCTA 2032
621ThrThrValGlnAlaAspGlys 628
2033 GATACAGTGTAAAGTG.....CGGGCTCTGGGAGGAGTT..... 2067
628 exTrpSerLeuThrValProAlaAlaMetGlyAsnLeuProAspGly 644
2068AA 2069
645 AlaValAlaIleThrAlaSerValThrAspLeuSerGlyAsnThrGlyAs 661
2070 CGCAGCGACGAGGAGTGATACCCAGAGTGGAGCAGTGTACATA. 2118
661 nThrSerArgThrIleThrValAspSerGlnAlaProAlaLeuSerIleA 678
2119 ..CCTGGCTGGATGAGAT.....GATGAATAACAATGGAATCCACCA 2160
678 spProLeuThrAlaAspAsnIleAsnAlaAlaGluSerGlyGlnAsp 694
2161 AGACCTGAAATTAATAGGATGATGTTCAACACAGCAAGTGTGTTTCAG 2210
695 LeuProIleThrGlyThrThrAspAlaGlnProGlyGlnThr...ValTh 710
2211 CAGAACATCCTCGGGAGGCTCATTTGCTGCTTCTGATGTCCTCAATGCTC 2260
710 rValThrLeuAsnGlyGlnThrTyrglnGlyValValGlnProAspGlyT 727
2261 CCATACCTGATCTCTCCACCTGGCCCAATCATCCGACCTGAAGCGGAA 2310
727 hrTrpSerValThrValProAlaAlaAsnVal..... 737
2311 ATTCAGGGGGCAGTCTCATT.....ATCTGACTTGGACAGCTCCTCG 2354
738GlyAlaLeuAlaAspGlyAsnAlaThrValThrAlaSerVa 751
2355 GGATGATTATGACCATGGACACTCACAAGTATATCATTCGAATAAGTA 2404
751 lAsnAspValAlaGlyAsnProser.....Serv 761
2405 CAAGTATCTTGTATCTCAGACAGCAAGTTCATGAATCTCTCAAGTGAAT 2454
761 alSerArgValAlaLeuValAspAlaThrProProValValThrIleAsn 777
2455 ACTACTCTCTC.....ATCCCAAGGAAGCAACTCTGA 2489
778 ProValAlaThrAspAsnValIleAsnThrProGluHisAlaGlnAlaGI 794

2490 GGAAGTCTTTTGTAAACACGAAAAACATTACTTTTGAATAATGCCACAG 2539
794 nIleIle.....SerGlyThrValThrGlyAlaGlnAlaGlyA 807
2540 ATCTTTTCATTCCTATTACAGGCTGTGATAGTGCATCTGAAATCAGAA 2589
807 sp.....IleValThrValThrLeuAsnValAspTyrThrVal 821
2590 ATATCCACACATTCGACGAGTATCTTTTGTATTCTCT..... 2625
822 ValAspGlySerGlyAsnTrpSerLeuGlyValProAlaSerValValSe 838
2625CCACAGACTCCGCCAGACAGACACCTAGTC 2653
838 rGlyLeuAlaAspGlySerTyrProValSerValSerValThrAspLysA 855
2654 CTGATGAAACGCTGCTCTCTGCTCTTAATATTCATATCAACAGACCCATT 2703
855 laGlyAsnThrGlySerGlnSerLeuThrValThrValAsnThrAlaAla 871
2704 ..CCTGGCATTCCACATTTTAAATTT 2727
872 ProLeuIleGlyIleAsnSerIle 879
seq_name: pir2.A69222

seq_documentation_block:
hypothetical protein MTH911 - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: A69222
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.
; Qiu, J.; Spadafora, R.; Vicaiere, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A:Reference number: A69000; MUID:98037514
A:Accession: A69222
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-698 <MTH>
A:Cross-references: GB:AE000866; GB:AE000666; NID:g2621998; PIDN:AAB85409.1; PID:g262
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH911

alignment_scores:
Quality: 136.00 Length: 627
Ratio: 0.437 Gaps: 28
Percent Similarity: 49.601 Percent Identity: 20.255
alignment_block:
US-09-949-696-20 x A69222 ..
Align seg 1/1 to: A69222 from: 1 to: 698
505 CTACGATGGGAGTATT.....GA 524
1 MetLysTrpGlyPheIleLeuPhePheGlyMetThrLeuLeuSerGlySe 17
525 CGAGTACAAATATGATGAGAAATTTCTACTTATCCCAATGGAAGAATACAAG 574
17 rSerTyrGlyAlaAsp.....TyrIleValAspAsnThrThrTyrLeuG 32
575 CAGTAAAGATGTCACGAGTATTACTGTACAAATGTAGTAAAGAAGTGT 624
32 lValPheThrProGlyGlyValGlnGlnGlyThrPheThrLeu..... 45
625 CAGGAGGACGCTGTACACCAAGA.....TGCATTTCAATAAGT 668
43 ...AsnGlyThrGluThrThrLeuGluAspGlyAspThrPheThrPheLe 61

[illegible]

332 gtyrLeuAspSerPheAsnAlaAraGsnGlyIleTyrSerSerGlyL 349
 1538 CA.....GTGACTGTGACACGACCGTGGGAAGGACACT 1572
 349 ysAsnAlaSerPhePheTyrIleThrSerAlaSerTyrAsnGlyGluAsn 365
 1573 TTGTTTCTTATCACTGGCAACACGACGCTCCCAAACTCTTCTCTGGGA 1622
 366 GlyIleLeuSerThrGlyProSerAlaSerMetGlnIleIleLeu...As 381
 1632 TCCAGTGTGACAGAAGCAAGTGGCTTTGTAGTGGCAAAAACACCAAAA 1672
 381 pValThrSerAsnSerGlnAsnGly.....ValSerArgGlyAspA 396
 1673 TGGCTACTCTCCAATCCACGAGCAATGCTAAGGTGGTGCACCTTGGAAATAC 1722
 396 snAlaTyrIleTrpPheValGluValLysLysAsnGly.....MetAsn 410
 1723 AGCTGTCAAGCAAGCTCAAAACCTTGACCTGACTGTACGTCCCGTGC 1772
 411 GlyIleHisSerSerGlyLysAsnLeuTyrLeuSerTyrValLysAsnAl 427
 1773 GTCCAATGCTACCCTGCCTCCCAATTACA..... 1800
 427 aThrAsnAsnThrLeuSerGlyIleAsnSerThrGlyIleAsnAlaArgI 444
 1801GTGACTTCCAAAACGAACGACCAAGGACACCAAAATTCGCCAGC 1842
 444 leLeuAspValThrValAsnLeuAsnArgGlyAsnGlyIleTyrSerSer 460
 1843CCTCGGTAGTTTATGCAAAAT..... 1863
 461 GlyTyrAsnThrThrIleAlaTyrPheGluAlaValAsnAsnArgGlyAs 477
 1864ATTGCCAAGAGCGCTCCCAATTTCTCAGGCGCAGTGTCA 1903
 477 nGlyLeuTrpIleSerGlyGlyArgSerTyrIleValGluGlyAsnThr 494
 1904 CA.....GCCTGATT 1914
 494 hrSerAsnArgGlnAsnGlyValArgValAsnGlyProAspAlaIleVal 510
 1915 GAATCAGTGAATGGAAAAACAGTTACCTTGGAACTACTGATATGGAGC 1964
 511 ArgSerValAsnAla.....ThrAspAsnThrAl 520
 1965 AGGTGCTGATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCAAA 2014
 520 aSerGlyIleAla.....IleTyrGlyLysAsnAlaIleI 532
 2015 CTTAT.....GACAGATGCTAGATACAGTGTA... 2043
 532 leTyrAsnCysThrSerMetArgAsnThrAspGlyValTyrGlyIleAla 548
 2044AAAGTCGGCGCTCTGGGAGGAT 2066
 545 SerPheArgMetIleLeuSerArgThrSerAlaAsnArgAsnCysGlyIle 565
 2067 TTAACGACGCCAGACGAGGAGTGATACCCAG 2097
 565 lAsnAlaLysGlyThrAlaThrIleGluGln 575

seq_name: pir2:C41859

```
seq_documentation_block:
```

seq_documentation_block;
IgaA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae
C; Species: Haemophilus influenzae

A; Variety: strain HK613

C:\Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C;Accession: C41859

R; Poulsen, K.; Reinholdt, J.; Kilian, A. *Reproductive Biology and Ecology* 174: 2012-2021. 1002

A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty
 A:Reference number: A41859; MUID:92234949
 A:Accession: C41859
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-1849 <POD>
 A:Experimental source: strain HK613
 A>Note: sequence extracted from NCBI backbone (NCBIP:97285)
 C:Superfamily: IGA-specific metalloendopeptidase
 C:Keywords: hydrolase; metalloproteinase

```

alignment_scores:
  Quality: 136.00      Length: 956
  Ratio: 0.346        Gaps: 53
  Percent Similarity: 41.109  Percent Identity: 21.339

alignment_block:
US-09-049-696-20 x C41859 ..

Align seg 1/1 to: C41859 from: 1 to: 1849

349 TCTACTCTCTCCAGGT.....AATGATGAACCTACACTGAGCAG..... 387
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
279 SerAspProLysGlyLeuSerGlnAspProLeuThrAsnTyrAlaVala 295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
388 .ATGGCAACTCTGGA.....GAGAAGGGTG 412
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 lIeuGlyAspSerGlySerProLeuPheValTyrAspArgGluLysGlyL 312
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
413 AAAGGATCACCTC...ACTCTGATTTTCATTCAGGAGAAAGTTAGCT 459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 ysrpLeuPheLeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLys 328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
460 GAATATGGACCACAAAGTAGGCAATTTGCCATGAGTGGGCTCATCTAGC 509
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
329 SerTrp...GlnGluTrpAsnIleTyrLysHisGluPheAla..... 341
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
510 ATGGGGAGTATTGACGAGGTACAAATAATGATGAGAAATCTACTTATCCA 559
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342 .....GluLysIleTyr..... 345

560 ATGGAGATACACAGCATGATGTTACAGCAGT...ATTACTGTGACA 606
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
346 .....GlnGlnTyrSerAlaGlySerLeuThrGlySer 356
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
607 AATGTAGTAAAGAAGTCTCAGGAGGAGCAGTGTACACCAAAAGATGCAC 656
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
357 Asn.....ThrGlnTyrThrTrpGlnAlaTh 365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
657 A.....TTCAATAAAGTAAACAGGACTCTATGAAAGAGGATGAGTTTG 700
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
365 rGlySerThrSerThrIleThrGly..... 373
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
701 TTCTCCAAATCCCGCCACAGCAGGAGGCTTCTATATGTTGTCACAAACAT 750
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
374 .....GlyGlyGluProLeuSerValAspLeuThrAspGly 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
751 GTTGATTCTATAGTAAATCTGTACAGCAACAAACACCAACAAAGC 800
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
386 LysAsp.....LysProAsnHisGlyLysSerI 395
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
801 TCCAAACACGAAATCAAAATGCATCTCCGA..... 834
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
395 eThrLeuLysGlySerGlyThrLeuThrLeuAsnAsnHisIleAspGlnG 412
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
834 ..... 834

412 lYAlaGlyGlyLeuPhePheGluGlyAspTyrGluValLysGlyThrSer 428
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
835 .....AGCACATGGGAAGTATCCGTGATCTTCGAGCAGCTTTAAGAAAC 878
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
429 AspSerThrThrTrpLysGlyAlaGlyValSerValAlaAspGlyLysTh 445
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

879 CACTCTATGACACACACAGCCACCAATCCCACTTCTCA...TTGCTGC 925
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
445 r.....ValThrTriPlysValHisAsnProLysTyrAspArgLeuAlaL 460
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
926 AGATTGGACAAGAATTGTGTGTAGTCTTGCAC.....AAA 963
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
460 ysrleGlyLysGly.....ThrLeuValGluGlyLysGlyLysAsn 474
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
964 TCTGGGAAGCATGCGACTGGTAACCCCTCAATCGACTGAATCAACAGCAGG 1013
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
475 GluGlyLeuLeuLysValGlyAsp.....G 483
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1014 CCAGCTTTCTCTCTGCAG.....ACAGTTGAGCTGG 1045
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
483 yThrValIleLeuLysGlnLysAlaAspAlaAsnAsnLysValGlnAlaP 500
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1046 GGTCTGGGTGGATGGTGCACATTTGCACAGTCTGCCCATGTACAAAGT 1095
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
509 heSerGlnValGlyIleValSerGlyArgSerThr..... 511
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1096 GAATCATACAGATAAAGCTGGCAGTGACAGGACACACACTCCCAAAAG 1145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
512 .....LeuValLeuAsn.....AspAspLysGlnValAspProAsnse 524
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1145 ATTACCTGCAGCAGCTTCAGAGGAGGAGCTCCATCTGCAGCGGCTTCGAT 1195
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
524 rIleTyrPheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnSerL 541
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1196 CGGATTACTGTCTAGGAAGAAATATCCAACTGATGGATCTGAAATT 1245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
541 euThrPheAspHisIleArgAsn.....IleAspAspGlyAlaArgVal 555
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1245 GTG.....CTGCTGAGGATGGGAAGACACACTATAAGTGGG..... 1284
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
556 ValAsnHisAsnMetThrAsnThrSerAsnIleThrIleThrGlyGluSe 572
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1284 ..... 1284

572 rLeuIleThrAsnProAsnThrIleThrSerTyrAsnIleGluAlaGlnA 589
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1285 .....TGC 1287

589 spAspAspHisProLeuArgIleArgSerIleProTyrArgGlnLeuTyr 605
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1288 TTTAACAGAGCTCAACAAAGTGTGCCATCATCCACACAGTCGCTTTGGG 1337
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
606 PheAsnGlnAspAsnArgSer.....TyrTyrThrLeuLysLysGly 619
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1338 GCCCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAAATGACA..... 1380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
619 yAlaSerThrArgSerGluLeuProGlnAsnSerGlyGluSerAsnGluA 636
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1381 .....GGAGGTTTACAGACATATGCTTCAGATCAAGTTTCAG 1416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
636 snTriPLeuTyrMetGlyArgThrSerAspGluAlaLysArgAsnValMet 652
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1417 AAC.....AATGGCTCTATTGATGCTTTTGGGGC 1445
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
653 AsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrPheGly.. 668
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1446 CTTTTATCAGGAATAGGAGTGTCTCTCAGCGCTCCATCCAGCTTGAGA 1495
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
669 ....GluGluGluThrLysAlaThrGlnAsnGlyLysLeuAsnValThrP 684
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1496 GTAAGGGATTACCTCCAGACACAGCCAGTGGATGAATGGCAGCATGATC 1545
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
684 heAsnGlyLysSerAspGlnAsnArgPheLeuLeuThrGlyGlyThrAsn 700
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1546 GTG.....GACAGCAGCTGGGAAAGGACACTTTTCTTATCACCTG 1589
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
701 LeuAsnGlyAspLeuAsnValGluLysGlyThrLeuPheLeuSerGlyAr 717
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

alignment_scores:
  Quality: 135.50      Length: 1008
  Ratio: 0.320         Gaps: 54
  Percent Similarity: 42.063      Percent Identity: 20.139

alignment_block:
US-09_049-696-20 x JC5575 ..

Align seg 1/1 to: JC5575 from: 1 to: 946

      82 CTGAGTAATTCACTCATTCAG.....CTGAACAACAATGGCTATGAAGG 125
      ||| |||||.....

```

945 GTGTTTAGTCCCTGCACAAATCTCGAAGCATTGGCGACTGTGTAAACCGCCTCA 994
946 :|||:|||::||| ||||||||| ||||||||| ||||||||| |||||||||
311 eLeuPheValIleAspValSerGlySerMet...TrpGlyIleLysMetU 327
995 ATCGACTGAATCAAGCAGGCCAGCTTTTCCTG....CTGCAGACAGTT 1038
996 :|||:|||::||| ||||||||| ||||||||| ||||||||| |||||||||
327 ySgInThrValGluAlaMetLysThrIleLeuAspAspLeuArgThrGlu 343
1039 GAGCTGGGCTCGTGTTGGATGGTGCACATTTGACAGTCTCTGCCCATGT 1088
1040 :||| ||| ::|||::||| ||||||||| ||||||||| |||||||||
344 AspGlnPheSer.....ValValAspPheAsnHisAsnValArgTh 357
1089 A...CAAAGTGAATCATACAGATAAACACTGCGCATGCGAGGACACAC 1135
1090 :|||:|||::||| ||||||||| ||||||||| ||||||||| |||||||||
357 rTrpArgAsnAspLeuValSerAlaThrLysThrGlnIleThrAsp.... 372
1136 TCGCCAAAAGNTTA.....CCTGCAGCAGCTTCAGAGGAGCGTCCATC 1179
1137 :|||:|||::||| ||||||||| ||||||||| ||||||||| |||||||||
373 AlaLysArgTyrlleGluLysIleGlnProSerGlyGlyThrAsnIle 398
1180 TGCAGCGGGCTTCGATCGGCATTTACTGTGATTAGGAAGAAA..... 1221
1181 :|||:|||::||| ||||||||| ||||||||| ||||||||| |||||||||
389 AsnGluAlaLeuLeuAlaIlePheIleLeuAsnGluAlaSerAsnLe 405
1222TATCCAACATGATGATCGAAATGTCTCTCAGCGGATG 1261
1223 :|||:|||::||| ||||||||| ||||||||| ||||||||| |||||||||
405 uGlyMetLeuAsnProAspSerValSerLeuIleValLeuValSerAspG 422
1262 GGGAGGACAACACATATAAGTGGGTCTTTAACGAGGTCAACAAGAAGTGT 1311
1263 :|||:|||::||| ||||||||| ||||||||| ||||||||| |||||||||
422 Ly...AspProThrVal.....GlyGluLeuLys..... 430
1312 GCCATATCCACACAGTCGCTTTGGGGCCCTCTCTCAGCTCAAGAACATGA 1361
430 430
1362 GGAGCTGTCCAAATATGACAGAGGTTTACAGACATATGCTTCAGATCAAG 1411
431LeuSerLys.....IleGlnLysAsnValLysGlnAsnI 442
1412 TTCGAACAANTGGCCTCATTCATGCTTTTGGG..... 1443
442 IeGlnAspAsnIleSerLeuPheSerLeuGlyIleGlyPheAspValAsp 458
1444GCCCTTTCATCAGGAATGGAGCTGTCTCTCAGCG 1478
459 TyrAspPheLeuLysArgLeuSerAsnGluAsnArgGlyIleAlaGlnAr 475
1479 CTCATCAGCTTGGAGCTAAGGGATTAAACCCTCCAGAACACCGCATGGGA 1528
475 gile.....TyrGlyAsnArgAspThrSerSerGlnLeuLysLysPheT 490
1529 TGAATGGCACAGTATCTGTGCAGACACCGGTGGGAAGGACACACTTGTGT 1578
490 yrAsnGln.....ValSerThrProLeuLeuArgAsnValIeInPhe 503
1573 CTTATCAGCTGGACAACGAGCAGCTCCCCAAATCCTCTCTGGGATCCCAG 1628
504 :|||:|||::||| ||||||||| ||||||||| ||||||||| |||||||||
1629 TGGACAGAAAGCAAGTGGCTTGTGTAGTGACAAACACCAAAATGGCCT 1678
513ValThrGlnAsnSerPheHisAsnT 521
1679 ACCTCCAAATCCAGGCATTGTAAGTTGGCAGCTTGGAAATACAGACTGTG 1728
521 yPheGlyClySerGluIleValValAlaGlyLysTyraSpProSerLys 537
1729 CAAGCAAGCTCAACAACCTTGACCTGACTGTCAAGTCCCGTGGCGTCCAA 1778
538 LeuAlaGluValGlySerIle...IleThrAlaThrSerThrAsnThrGlu 553

```
1779 TGCTACCTGCTCCCAATTACA.....GTGACTT 1807
553 uLeuValLeuGluThrLeuSerGlnMetAspLeuGluAspPheLeuS 570
1808 CCAAAACGAAACAGGACACAGCAATTCGCCAGCCCTCTGGTAGTTAT 1857
570 erLysAspLysHisAlaAspProAsnPheThrLysLysLeuTrpAlaTyr 586
1858 GCAAAATTGCGCAA.....GGAGCCTCCCAATTCTCAG 1892
587 LeuThrIleAsnGlnLeuLeuAlaGluArgSerLeuAlaPro..... 600
1893 GGCCAGTGTACAGCCCTGATTGAATCAGTGAATGGAAACACTT.... 1938
601 .....ThrAlaAlaIleLysArgLysIleThrLysThrIleLeuG 614
1939 .....ACCTTGGAACTACTG 1953
614 lnMetSerLeuAspHisHisIleValThrProLeuThrAlaMetValIle 630
1954 GATAATGGAGCAGGTGCTGATGCTACTAGGATGACGGTGTCTACTCAAG 2003
631 GluAsnGluAlaGlyAspGluArgMetLeuAlaAspSerPro..... 644
2004 GTATTTCAACAATTATGACACAGAAATGGTAGATACAGTGTAAAGTGGGG 2053
644 ..... 644
2054 CTCTGGGAGGAGTTAACGCAGCAGGAGAGTGTATACCCACAGCAG... 2100
645 .....ProGlnAspHis 648
2101 .....AGTGAGCACTGTAC..... 2115
649 SerCysSerGlyAlaLeuTyrTyrglyThrLysValAlaSerAlase 665
2116 ATACCTGGCTGG..... 2127
665 rIleProSerTrpAlaSerProSerProThrProValMetAlaMetLeuA 682
2128 ..ATTGAGAATGATGAATACAATGAATCCCAAGACCTGAATTAAT 2175
682 laValGlyAlaAsnArgLeuGluSerThrProProHisValIleArg 698
2176 AAGGATGATGTTCAACAC.....AAGCA 2198
699 vaGluAsnAspProHisPheIleIleTyrLeuProLysSerGlnLysAs 715
2199 AGTGTGTTTCAGCAGAACATCCTCGGGAGGCTCATTTGTGGCTTCTCATG 2248
715 nIleCysPheAsnIleAspSerGlu..... 723
2249 TCCCAAAATGCTCCCATACCTGATCTCTTCCCACTGGCCAAATCACCGAC 2298
724 .....ProGlyLysIleLeuSer 729
2299 CTGAGGCGGAATTCACGGGGCAGTCTCATTAATCTGACTTGGACAGC 2348
730 LeuValSerAspProGluSerGlyIleLeuValAsn....GlyGlnLeuIl 745
2349 TCCTGGGGATGATTATGACCATGCAACAGCTCATATATC..... 2391
745 eGlyAlaLysLysAlaGluAsnGlyLysLeuArgThrTyrPheGlyLysL 762
2392 .....ATTGCAATTAAGTACAAGT 2409
762 euGlyPheTyrPheGlnLysGluAspMetLysIleGluIleSerThrGlu 778
2410 ATTCTTGATCTCAGACACAGTTCATGAATCTCTTCAAGTGAATACTAC 2459
779 AsnIleThrLeu.....IleAsnGlySer.....SerThrTh 789
2460 TGCTCTCATCCCAAGGAAGCC.....AACTCTGAGGAAGTCTTTT 2500
```

```
1779 rSerLeuPheTrpSerAspThrAlaHisLeuGlyAsnGlnArgValLeuI 806
2501 TGTTT.....AAACCAAAACACTTACTTTTGAATAATGCCACAGATCTT 2544
806 leSerValLysLysGlyLysSerValThrLeuThrLeuAsnLysGluMet 822
2545 TTCATTGCTATTTCAGGCTGTTGATAGGTGATCTGAAATCAGAAATATC 2594
823 PhePheSerVal...LeuLeuHisHisValTrpLysLysHisProVal.. 837
2595 CAACATTGCACGAGTATCTTTGTTTATTCTCCACACACTCCGCCACAGA 2644
838 ..AsnValAspPheLeuGlyIleTyrIleuProPro...ThrAsnLysPheS 853
2645 CACTAGTCTCTGATGAACGCTCTGCTCTTGT.....CCTAATATT 2685
853 erProSerAlaHisGlyLeuLeuGlylnPheMetAsnLysProAsnIle 869
2686 CATATCAACAGCACCATTCCTGGC 2709
870 HisIlePheAsnGluArgProGly 877
```

us-09-049-696-20.rpr

Tue Apr 2 09:39:58 2002

This Page Blank (uspto)

Tue Apr 2 09:39:58 2002

474 AGGTAGGGCATTTGTCCTATGAG.....TGGGCTC 502
:::|:::|:::|:::|:::|:::|
469 pGlyAlaasnPheThrTySerLeuGlnAspAlaLeuThrGlyLeuThrS 486
503 ATCTACGATGGGGAGTATTGACGAGTACAATAATGATCGAATAATCTAC 552
::|::|::|::|::|::|::|::|::|::|::|::|::|
486 erileThrLeuGlyGlyThrThrAsnGlyGlyAsnAspAlaLysThrVal 502
553 TTATCC...AATGGAAGAATACAACGAGTAAGATCTTCAGCAGAGTATTAC 599
:::|::|::|::|::|::|::|::|::|::|::|::|::|
503 IleAsnLysAspGlyLeuThrIleThrProAlaGlyAsnGlyGlyThrTh 519
600 TGCTACAAATGTAGTAAAGAAATGTCAGGGAGGCAGCTGTTACACCAAA 649
::|::|::|::|::|::|::|::|::|::|::|::|::|
519 rGlyThrAsnThrIleSerValThrLysAspGlyIleLysAlaGlyAsnL 536
650 GATGCCACATTCATAAGTAACAGCACTC....TATGAAAAAGGATGT 693
:::|::|::|::|::|::|::|::|::|::|::|::|::|
536 ysAlaIleThrAsnValAlaSerGlyLeuArgAlaTyTraspAlaAsn 552
694 GAGTTTTCTTCCAATCCCGCCAGCAGAGAGAGGCTTCTATAATCTTGC 743
:::|::|::|::|::|::|::|::|::|::|::|::|::|
553 PheAspValLeuAsnSerAlaThrAsp.....LeuAs 564
744 ACAACATGTTGATCTATAGTTCAATTCGTACAGAACAAACCACACA 793
:::|::|::|::|::|::|::|::|::|::|::|::|::|
564 nArgHisValGluAspAlaTyTrLysGlyLeuLeu.....AsnLeuAsnG 579
794 AAGAAGCTCCAAACACAGCAAATCAAAAATGCAATCTCGGAAGCACATGG 843
:::|::|::|::|::|::|::|::|::|::|::|::|::|
579 LuLysAsnAlaAsnLysGln.....585
844 GAATGATGCCGTGATCTGAGGACTTTAGAAACCACCTCTCATGCACAC 893
:::|::|::|::|::|::|::|::|::|::|::|::|::|
586ProLeuValTh 589
894 ACAGCCACCAATCCCACCTCTCATTTGCTGCAGATTGGACAAAGAATTG 943
:::|::|::|::|::|::|::|::|::|::|::|::|::|
589 raspSerThrAlaAlaThrValGlyAspLeu.....ArgLysLeuG 603
944 TGTGTTTAGTCCTTGACAATCTGGAAGCAGTCGGGACTGGTAACCCCTC 993
:::|::|::|::|::|::|::|::|::|::|::|::|::|
603 lyTrpValValSerThrLysAsnGly.....ThrLysGluLysU 616
994 RATCGACATGAATCAAGCAGCGCCAGCTTTCTGCTGCACACAGCTTGAGCT 1043
:::|::|::|::|::|::|::|::|::|::|::|::|::|
617 AsnGlnValLysGlnAlaAspGluVal.....Le 626
1044 GGGGTCTGGGTGGGATGGTGACATTTGACAGTGTGCCATGTCACAAA 1093
::|::|::|::|::|::|::|::|::|::|::|::|::|
626 upheThrGlyAlaGlyAlaAlaThrValThrSerLysSerGluAsnGlyL 643
1094 GTCACTCATCAGATAAACAAGTGGC.....1119
yhisThrIleThrValSerValAlaGluThrLysAlaAspCysGlyLeu 659
1120 ...AQTGACGGGACACACTCCCCAAAAGATTACT.....1152
:::|::|::|::|::|::|::|::|::|::|::|::|::|
660 GluLysAspGlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAs 676
1153GCACAGCTTCAGGGAGGACGCTCCATCTGACAGGGGGCTTC 1192
:::|::|::|::|::|::|::|::|::|::|::|::|::|
676 pAsnValLeuThrValGlyAsnAsnGlyThrAlaValThrLysGly.... 691
1193 GATCGGCATTTACTGTGATTAGGAAGAAATATCCAACCTGATGGATCTGAA 1242
:::|::|::|::|::|::|::|::|::|::|::|::|::|
692GlyPheGluThrVal...LysThrGlyAlaThrAspAlaAspArg 705
1243 ATTGTCTCTCTCAGCGATGGGGAACACACACTATAGCTGGGTCTTTAA 1292
:::|::|::|::|::|::|::|::|::|::|::|::|::|
706GlyLysValThrValLysAspAlaThrAlaAs 716
1293 CGAGGTCAAAACAAGTGTGCCATCATCCACACAGCTGCTTTGGGCGCT 1342

[illegible]

```
986 nLysSerGlyTrpArgValThrGlyGluGlyAlaThrAlaGluThrGlyA 1003
2062 .....GGAGTTAAACGACGACGAGAGTGTATACCCAGCAGAGTGA 2106
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1003 laThrAlaValAsnAlaGlyAsnAlaGluThrValThrSerGlyThrSer 1019
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2107 GCACGTGTACATACCTGGCTGGATTGAGAAATGATGAAATACATGAATCC 2156
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1020 ValAsnPheLysAsnGly.....AsnAl 1027
2157 ACCAAGACCTGAATTAATAAGGAT.....G 2182
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1027 aThrThrAlaThrValSerLysAspAsnGlyAsnIleAsnValLysTyra 1044
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2183 ATGTT.....CAACACAGCAAGTGTG 2205
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1044 spValAsnValGlyAspGlyLeuLysIleGlyAspAspLysIleVal 1060
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2206 TTCACGACAGACATCC.....TCGGAGGCTCATTTGTG..... 2238
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1061 AlaAspThrThrLeuThrValThrGlyGlyLysValSerValProAl 1077
2239 ....GCTTCTGTATGCCAAATGCTCCCATACCTGATCTCTCCACCTG 2284
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1077 aGlyAlaAsnSerValAsn..... 1084
2285 GCCAATACCCAGCTGAAGGGGGAATTCACGGGGCGAGTCTCATTAAT 2334
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1085 ..AsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsnAsn 1100
2335 CAGCTTGGACAGCTCCCTGGGATGATGATGACCATGGAACAGCTCACAA 2384
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1101 LeuSerTrpThrAlaLysAlaAspLysTyraAlaAspGlyGluSerGlu 1117
2385 GTATATC..... 2391
1117 yGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysAlaG 1134
2392 .....ATTCGATAAGTACAGATTTCTTGATCTCAGACAGATTC 2433
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1134 lLysAsnLysValLysGlnSerGluLysAspPheThrTyrSerLeu 1150
2434 AATGAATCTTCAAGTGAATACTGCTCTCATCCCAAGGACCA 2483
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1151 GlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAlaAs 1167
2484 C.....TCTGAGAGTCTTTTGTAAACAGAAACATTACTTTT 2526
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1167 nGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrIle 1184
2527 .....GAAATGGCACAGAT 2541
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1184 hrLeuAlaAsnGlyAlaAlaLaglyThrAspAlaSerAsnGlyAsnThr 1200
2542 CTTTCATTGCTATTACAGCTGTTGATAA.GTCGATCTGAATCAGAAA 2590
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1201 lIleSerValThrLysAspGlyIleSerAlaGlyAsnLysGluIleThrAs 1217
2591 TATCCAACATTGCACGAGTATCTTTTATTCTCCACAGACTCCGCCA 2640
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1217 nValLys.....SerAlaLeuLysThrTyrLysAspThrGlnAsnThrA 1232
2641 GAGACACCTAGTCTGATGAAGGCTGCTCTTGTCTTAATATTCATAT 2690
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1232 laAspGlu..... 1234
2691 CAACGACCACTTCTGGCATTACATTTTAAATAATTATGTGAAGTGA 2740
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1235 ThrGlnAspGluPheIleAlaValLysAsnAlaAsnGluValG 1251
2741 T 2741
1251 u 1251
```

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-685-467-4

seq_documentation_block:

Sequence 4, Application block US/08685467

Patent No. 6060059

GENERAL INFORMATION:

APPLICANT: St. Geme III, Joseph W.

APPLICANT: Barenkamp, Stephen J.

TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/685,467

FILING DATE: 22-JUL-1996

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/409,995

FILING DATE: 24-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Silva, Robin M.

REGISTRATION NUMBER: 38,304

REFERENCE/DOCKET NUMBER: A-61053-2/RET/RMS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1912 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-685-467-4

alignment_scores:

Quality: 141.50

Ratio: 0.321

Percent Similarity: 45.605

Length: 967

Gaps: 44

Percent Identity: 18.718

alignment_block:

US-09-049-696-20 x US-08-685-467-4 ..

Align seg 1/1 to: US-08-685-467-4 from: 1 to: 1912

274 ACATGGAAGACAAAGCTGACTATGTGAGACCAAAACTTGAGACCTACAA 323

428 SerTyrLysAlaLysAlaGlu..... 434

324 AATGCTGATGTTCTGTTGCTGAGTCTACTCTCCAGGTAATGATGAAC 373

435AlaAsp..... 444

374 CCTACACTGAGCAGATGGCAACTGTGAGAGAGGTTGAAAGGATCCAC 423

444 lSerLysAspGlnGluValLysAlaGly..... 454

424 CTCACCTCTGATTTTCATTTCAGGAAAAAAGTTAGCTGAATATGGACACA 473

455 ...ThrValThrPheLysAlaGlyLysAsnLeu...LysValLysGlnAs 469

[illegible]

1293 CGAGGTCAACAAGAGTGTGGCCATCATCCACACAGTCGCTTTGGGGCCCT 1342
716 nspAlaAspLysLeuValAlaThrValLysAspValAlaThrAlaThr 733
1343 CTCACGCTCAAGAACTAGAGGAGCTGTCCAAAATGACAGGAGGTTTACAG 1392
733 snSerAlaAlaThrPheValLysThrGluAsnLeuThrThrSerIleAsp 749
1393 ACATATGCTTCAGATCAAGTTCAGAACAAATGGCCCTCATGTGCTTTGG 1442
750GluAspAsnProThrAspAsnGlyLysAspAsp..... 760
1443 GGCCCTTTCATCAGAAAT.....GGAGCTGCTCTCTCAGCGCT 1480
761 AlaLeuLysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeu 777
1481 CCATCAGCTTGAGAGTAGGATTAAACCTC.....CAGAACACG 1521
777 ysValLysArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeu 793
1522 CAGTGGATCAATGGCACAGTATCGTGGACAGCACCGGTGGAAAGACAC 1571
794 GluValLysThrAlaLysValSerAspThrLeuThrIleGlyGlyAsnThr 810
1572 TTTGTTTCTTATCAGCTGCACACACGACCCCTCCCAATC..... 1611
810 r.....ProThrGlyGlyThrThrAlaThrProLysValAsnIleThr 825
1611CTCTCTGGGATCCAG 1628
825 erThrAlaAspGlyLeuAsnPheAlaLysGluThrAlaAspAlaSerGly 841
1612CTCTCTGGGATCCAG 1628
842 SerLysAsnValTyLeuLysGlyIleAlaThrThrLeuThrGluProSer 858
1629 TGGACAGAACGAAGTGGCTTTGTA.....GTGGACAAAACACCA 1669
858 rAlaGlyAlaLysSerSerHisValAspLeuAsnValAspAlaThrLysL 875
1670 AAATGSCCTACCTCAAAATCCAGGCATTGTCAAGTTGGCACCTGGAAA 1719
875 ysSerAsnAlaAlaSerIleGluAspValLeuArgAlaGly..... 888
1720 TACAGTCTCAAGCAAGCTCAACAACCTTGACCTGACTGTCAAGTCCCG 1769
889 TrpAsnIleGlnGlyAsnGlyAsnAsnValAspTyrValAlaThrTyrAs 905
1770 TCGGTCCCAATGCTACC.....CTGCCTCCAATTACACTGA 1804
905 pThrValAsnPheThrAspSerThrGlyThrThrValThrValThr 922
1805 CTTCACAAACGAACGAAGCACACACAGCAAAATTCCTCCAGCCCTCGTAGT 1854
922 hrGlnLysAlaAspGlyLysGly..... 929
1855 TATGCAATATTGCCCAAGGAGCTCTCCCAATTCCTCAGGGCCAGTGCAC 1904
930 ...AlaAspValLysIleGlyAlaLys.....Th 938
1905 AGCCCTGATTGAATCAGTGAATGGAAAACAGTTACC...TTGGAACCTAC 1951
938 rSerValIleLysAspHisAsnGlyLysLeuPheThrGlyLysAspLeu 955
1952 TGGATATGGACAGGTGCTGATGCTACTAAGGATGACGCT..... 1992
955 ysAspAlaAsnAsnGlyAlaThrValSerGluAspAspGlyLysAspThr 971
1993GTCTACTCAAGGTATTTCACACTTATGCACGAGTGG 2030
972 GlyThrGlyLeuValThrAlaLys.....ThrValIleAspAlaValAs 986
2031 TAGAVACAGTGTAAAGTCGCGGCTCTGGGA..... 2061

986 nlySerGlyTrpArgValThrGlyGluGlyAlaThrAlaGluThrGlyA 1003
2062CGAGTTAACCCAGCCAGACGGAGGTGATACCCACGACAGTGA 2106
1003 laThrAlaValAsnAlaGlyAsnAlaGluThrValThrSerGlyThrSer 1019
2107 GCATGTGTACATACCTGGCTGGATTGAGNATGATGAATACATGGAAATCC 2156
1020 ValAsnPheLysAsnGly..... 1027
2157 ACCAAGACCTGAATTAATAAGGAT.....G 2182
1027 aThrThrAlaThrValSerLysAspAsnGlyAsnIleAsnValLysTyrA 1044
2183 ATGTT.....CAACAACAAGCAAGTGTGT 2205
1044 spValAsnValGlyAspGlyLeuLysIleGlyAspAspLysLysIleVal 1060
2206 TTCACGACAGACATCC.....TCGGGAGGTCATTTGTG..... 2338
1061 AlaAspThrThrLeuThrValThrGlyGlyLysValSerValProAl 1077
2239GCTTCTGATGTCCTCAATAGTCTCCCATACCTGATCTCTCCACCTG 2284
1077 aglyAlaAsnSerValAsnAsn..... 1084
2285 GCCAATCACCGACCTGAAGCGGGAAATTCACGGGCGAGTCTCATTAAT 2334
1085 ..AsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsnAsn 1100
2335 CTGACTTGGACAGCTCTGGGATGATTATGACCATGGAACAGCTCAAA 2384
1101 LeuSerTprThrAlaLysAlaAspLysTyrAlaAspGlyGluSerGluGl 1117
2385 GTATATC..... 2391
1117 yGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysAlaG 1134
2392ATTCGAATAAGTACAAGTATTCGTGATCTCAGACACAAGTTC 2433
1134 LyLysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSerLeu 1150
2434 AATGAATCTTCAAGTGAATACTACTGCTCTCATCCCAAGGAAGCCAA 2483
1151 GlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAlaAs 1167
2484 C.....TCTCAGGAAGCTTTTGTGTTTAAACCAGAAAACATTAATTTT. 2526
1167 nGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrIle 1184
2527GAAATATGGCACAGAT 2541
1184 hrLeuAlaAsnGlyAlaAlaGlyThrAspAlaSerAsnGlyAsnThr 1200
2542 CTTTTCATTCATTCAGGCTGTTGATAA..GGTCGATCTCAAAATCAGAAA 2590
1201 ileSerValThrLysAspGlyIleSerAlaGlyAsnLysGluIleThrAs 1217
2591 TATCCAACATTGCACGAGTATCTTTGTTTATTCTCTCCACAGACTCCGCCA 2640
1217 nValLys.....SerAlaLeuLysThrTyrLysAspThrGlnAsnThrA 1232
2641 GAGACACCTAGTCTCTGATGAACCTGCTCTCTTGTCTCTAATATTCATAT 2690
1232 laAspGlu..... 1234
2691 CAACGACCATTCCTGGCATTCACATTTTAAAAATATATGTGAAGTGA 2740
1235 ThrGlnAspLysGluPheHisAlaValIlyAsnAlaAsnGluValGl 1251
2741 T 2741
;

```

1251 u 1251
seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-377-155-33
seq_documentation_block:
: Sequence 33, Application US/09377155
Patent No. 6197312
GENERAL INFORMATION:
APPLICANT: PEAK, Ian Richard Anselm
APPLICANT: JENNINGS, Michael Paul
APPLICANT: MOXON, E. Richard
TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
FILE REFERENCE: 065064/0128
CURRENT APPLICATION NUMBER: US/09/377,155
CURRENT FILING DATE: 1999-08-19
PRIOR APPLICATION NUMBER: PCT/AU98/01031
PRIOR FILING DATE: 1998-12-14
PRIOR APPLICATION NUMBER: GB 9726398.2
PRIOR FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0.
SEQ ID NO 33
LENGTH: 2353
TYPE: PRP
ORGANISM: Haemophilus influenzae
US-09-377-155-33

alignment_scores:
: Quality: 141.50 Length: 967
: Ratio: 0.321 Gaps: 44
Percent Similarity: 45.605 Percent Identity: 18.718

alignment_block:
US-09-049-696-20 x US-09-377-155-33 ..
Align seg 1/1 to: US-09-377-155-33 from: 1 to: 2353

274 ACATGGAAGCAAAAGGTGACTATGTGAGACCAAACTTGAGACCTACAA 323
:::|||||::::|||:::
429 SerTrpLysAlaLysAlaGlu..... 435
324 AAATGCTGATGTTCTGGTTCCTGAGTCTACTCCTCCAGGTATGATGAC 373
||||| ::::
436 ...AlaAsp.....ThrAspGlyAlaLeuGluGlyI 445
374 CCTACACTGAGCAGATGGCGCAACTGTGCAGAGAAGGGTCAAAGGATCAC 423
:::||||| ::::
445 leSerLysaspGlnGluValLysAlaGlyGlu..... 455
424 CTCACPTCCTGATTTCATTCGACGAAAAAAGTAGCTGAATATGGACCACA 473
||| ||| ||| ||| ||| ||| :::
456 ...ThrValThrPhelysAlaGlyLysasnLeu...LysvalLysGlnas 470
474 AGGTAGGCCATTTGTCCTCATGAG.....TGGGCTC 502
||| ||| ::::
470 pGlyAlaasnPheThrTyrSerLeuGlnAspAlaLeuThrGlyLeuThrs 487
503 ATCTACAGTGGGAGCATTTTCGACGAGTACAATAATGATGACAAAATTCAC 552
::: ||| ::: ||| ||| |||
487 exilethrLeuGlyGlyThrThrAsnGlyGlyAsnAspAlalysThrVal 503
553 TTATCC...AATGGAAGATACAAGCAGTAGTAAGTGTTCAGCAGGATTTAC 599
::: ||| :::
504 lleAsnLysaspGlyLeuThrIlethrProAlaGlyasnGlyGlyThrThr 520
600 TGGTACAAATGTAGTAAAGAAGTGTCTCAGGAGGAGCAGCTGTTACACCAAA 649
||||| ||| ::: |||
520 rGlyThrAsnThrIleSerValThrLysAspGlylleLysAlaGlyAsnL 537
650 GATGCACATTCATAAGTTAACAGGACTC.....TATGAAAAAGGATGT 693
::: ||| ::: ||| ||| |||
537 ysAlallethrAsnValIleSerGlyLeuArqAlaThrAspAlaAsn 553

```



```

1062 AlaAspThrThrLeuThrValThrGlyGlyLysValSerValProAl 1078
2239 ....GCTTCTGATGTCCTCCAAATGCTCCCATACCTGATCTCTCCACCTG 2284
1078 aglyAlaAsnSerValAsnAsn..... 1085
2285 GCCAATACCCAGCTGAGGGGGAATTCACGGGGCAGCTCTCATTAAT 2334
1086 ..AsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsnAsn 1101
2335 CTGACTTGACAGCTCTCTGGGATGATTATGACCATGGAGACAGCTCACAA 2384
1102 LeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyGluSerGluGl 1118
2385 GTATATC..... 2391
1118 yGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysAlaG 1135
2392 .....ATCGAATAAGTACAGTATTCTTGATCTCAGACAGTTC 2433
1135 LysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSerLeu 1151
2434 AATGAATCTCTCAAGTGAATCTACTGCTCTCATCCCAAGGAGCCAA 2483
1152 GlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAlaAs 1168
2484 C.....TCTGAGGAAGTCTTTTGGTTTAAACCAGAAACATTACTTTT. 2526
1168 nGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrIleT 1185
2527 .....GAAATGGCACAGAT 2541
1185 hrLeuAlaAsnGlyAlaAlaAlaGlyThrAspAlaSerAsnGlyAsnThr 1201
2542 CTTTTCATTGCTATTTCAGGCTGTGTATAA.GGTGATCTGAAATCAGAAA 2590
1202 IleSerValThrLysAspGlyIleSerAlaGlyAsnLysGluIleThrAs 1218
2591 TATCCAAATTCGACGAGTATCTTTGTTTAAACCAGAAACATTACTTTT. 2640
1218 nValLys.....SerAlaLeuLysThrTyrLysAspThrGlnAsnThrA 1233
2641 GAGACACCTAGTCTCTGATGAACGCTGCTCTCTGTCCTTAATATTCATAT 2690
1233 laAspGlu..... 1235
2691 CAACAGCACCATTCTGCGCATTCACATTTTAAAAATATGTGGAAGTGA 2740
1236 ThrGlnAspLysGluPheHisAlaAlaValLysAsnAlaAsnGluValGl 1252
2741 T 2741
1252 u 1252

```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:us-08-913-942-4

seq_documentation_block:
; Sequence 4, Application US/08913942
; Patent No. 6200578

GENERAL INFORMATION:
APPLICANT: St. Geme, Joseph
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr Hobbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,942
FILING DATE: 29-DEC-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/4031
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vance, Dolly A.
REGISTRATION NUMBER: 39,054
REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2353 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-913-942-4

alignment_scores:
Quality: 141.50 Length: 967
Ratio: 0.321 Gaps: 44
Percent Similarity: 45.605 Percent Identity: 18.718

alignment_block:
US-09-049-696-20 x US-08-913-942-4 ..

Align -seg 1/1 to: US-08-913-942-4 from: 1 to: 2353

```

274 ACATGGAAGACAAAGGCTGACTATGTGACCAACCAAACTTGAGACCTACAA 323
429 SerTpLysAlaLysAlaGlu..... 435
324 AAATGCTGATGTCTGTTGCTGAGTCTACTCTCCAGCTAATGATGAAC 373
436 ....AlaAsp.....ThrAspGlyAlaLeuGluGlyI 445
374 CCTACACTGACGAGATGGCAACTGTGGAGAGAAGGTGAAAGGATCCAC 423
445 leSerLysAspGlnGluValLysAlaGlyGlu..... 455
424 CTCACCTCTGATTTCATTCAGGAAAAAGTTAGCTGAATATGGACCACA 473
456 ...ThrValThrPheLysAlaGlyLysAsnLeu...LysValLysGlnAs 470
474 AGGTAGGGCATTGTGTCATGAG.....TGGGCTC 502
470 pGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuThrGlyLeuThrS 487
503 ATCTACGATGGGAGTATTTGAGGAGTACAATAATGATGAGAAATCTAC 552
487 erIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspAlaLysThrVal 503
553 TTATCC...AATGGAAGATACAAACAGTAAGATGTTTCAGCAGGTATTAC 599
504 IleAsnLysAspGlyLeuThrIleThrProAlaGlyAsnGlyGlyThrThr 520
600 TGGTACAAATGTAGTAAAGAGTGTACGGAGGCGAGCTGTTACACCAAAA 649
520 rGlyThrAsnThrIleSerValThrLysAspGlyIleLysAlaGlyAsnL 537

```

776 ysVallyArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeu 794
 :

1522 CAGTGGATGAATGCACAGTCATCGTGACAGCACCCGTGGAAAGCACC 1571
 :
795 GluValThyTrpAlaLysValSerAspThrLeuThrIleGlyAsnth 811
 :

1572 TTTGTCTTATTACCTCGACAACGCCAGCTCCCAATC..... 1611
 | :
811 F.....ProThrGlyGlyThrAlaThrProLysValAsnIleThrS 826
 :
1613 erThrAlaAaspGlyLeuAsnPheAlaLysGluthrAlaAaspAlaSerGly 842
 :
1612CTTCTCTGGATCCCAG 1628
 :
843 SerLysAsnValTyLeuLysGlyIleAlaThrThrLeuThrGliUprose 859
 :
1629 TGGACAGAAGCAAGGTGGCTTGTA.....GTGGCAAATAAACCCA 1669
 |:| :
859 rAlaGlyAlaySerHisValasLeuAsnValAspAlathrLysl 876
 :
1670 AAATGGGCTACTCCAAATCCCCAGGACTTGCTAAGGTGGCAGCTGGAAA 1719
 || :
876 ySerAsnAlaAserIleGluAspValLeuArgAlaGly..... 889
 :
1720 TACAGTCTGCAAGCAAGCTCAACAACCTTGACCGTGACTGTCAAGTCCCG 1769
 :
890 TrpAsnIleGlnGlyAsnGlyAsnValAspTryValAlatThrTyraS 906
 :
1770 TGCGTCCAAATGCTACC.....CTGCCCTCCAAATTACAGTAGTA 1804
 :
906 pThrValasnPhetrAspAspSerThrGlyThrThrValthrValt 923
 :
1805 CTTCAAAAAGCAACGAGCACACCAGCAAAATCCCGCCCTCTCGTAGTT 1854
 :
923 hrGlnLysAlaaSpGlyLysgly..... 930
 :
1855 TATGCAANTATTCGCAAGAGAGCTCCCAATTCTCAGGCCAGTGTCCAC 1904
 |||| :
931 ...AlaaSpvallysiilegilyAlalaYs.....th 939
 :
1905 AGCCCTGATTCAATCAGTGAATGGAATAACAGTATACC...TTGGAAGTAC 1951
 |:| :
939 rSerVallleylasyaspHistsangLyLysLeuPheThrGLylasyapLeuL 956
 ||| :
1952 TOGATATPGACGAGCGTCTCATCTACTAAGGATACGGT..... 1992
 ||| :
956 yaspaAlaaaSnnglyAlalatrValSerGluAspAspglyLysaspThr 972
 :
1993GTCTACTCAAGTATTTCACAACITATGACACGAGATGG 2030
 |||| :
973 GlyThrGlyLeuvAltThrAlaLYs.....ThrVailieaspAlavalAS 987
 :
2031 TAGATACAGTCTAAAGTGGCGGCTCTGGGA..... 2061
 :
987 nlysSerGlyTrpargValthrGlyGluGlyAlalathralagluthrglya 1004
 :
2062GGAGTTTAACGCCACGACGAGCGAGTGTATACCCECCAGAGATGGA 2106
 lathRAlavaLasnalaglyasnAlacgluthrValthrSerGlyThrSer 1020
 :
2107 GCATGTATACATACCTGGCTGGATTGAGAAATGANAATACATGGAAATCC 2156
 :
1021 ValasnPheLysasnGly.....ASNAL 1028
 :
2157 ACCAAGACCTCAAAATTAATAAGGAT.....G 2182
 :
1028 athrThrAlatrValserLysaspasnGlyAsnIleaSnVallystrya 1045
 :
2183 ATGTT.....CAACAACAAGCAAGTGTGT 2205
 |||| : : : : : : : : : : : : : : : :

1045 spValasnValGlyAspGlyLeuLysIleGlyAspAspLysLysIleVal 1061
2206 TTCAGCAGAACATCC.....TCGGGAGGCTCATTTGTG..... 2238
::: |||||:::
1062 AlaAspThrThrLeuThrValThrGlyLysValSerValProAl 1078
2239GCTTCGTGATGCCCAATGCTCCCATACCTGATCTCTCCACCTG 2284
|||||:::||||
1078 aGlyAlaAsnSerValAsnAsn..... 1085
2285 GCCAAATCACCGACTCGAAGCGGAATTCACGGGGCGAGCTCATTAAT 2334
::: |||||:::
1086 ..AsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsnAsn 1101
2335 CTGACTTGGACACTCTCTGGGGATGATTATGACCATGGACACTCACAA 2384
|||||:::||||
1102 LeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyLeuSerGlu 1118
2385 GTATATC..... 2391
1118 yGluThrAspGluValLysAlaGlyAspLysValThrPheLysAlaG 1135
2392ATTGGAATAGTACAGTATTCTTGATCTCAGACAAAGTTC 2433
::: |||||:::
1135 LysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSerLeu 1151
2434 AATGAATCTCTCAAGTGAATACTACTGCTCATCCCAAGAACGCCAA 2483
::: |||||:::
1152 GlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyThrAlaAs 1168
2484 C.....TCGAGGAAGTCTTTTGTGTTTAAACCAAGAAACATTACTTTT 2526
|::: |||||:::
1168 nGlyArgAsnAspThrGlyThrValLysLysAspGlyLeuThrIleT 1185
2527GAAATGGCACAGAT 2541
::: |||||:::
1185 hrLeuAlaAsnGlyAlaAlaGlyThrAspAlaSerAsnGlyAsnThr 1201
2542 CTTTTCATCTCTATCAGGCTGTGTATAA..GGTCGATCTGAAATCAGAA 2590
|::: |||||:::
1202 IleSerValThrLysAspGlyIleSerAlaGlyAsnLysGluIleThrAs 1218
2591 TATCCAAACATTGCACAGATATCTTTGTTTATTCTCCACAGACTCGGCCA 2640
|::: |||||:::
1218 nValLys.....SerAlaLeuLysThrTyrLysAspThrGlnAsnThrA 1233
2641 GAGACACCTAGTCTGATGAACGCTGCTGCTCTGCTCTTAATATTCATAT 2690
|||
1233 laAspGlu..... 1235
2691 CACACACCATCTCTGGCATTACATTTTAAATAATATGCGAAGTGA 2740
|||::: |||||:::
1236 ThrGlnAspLysGluPheHisAlaAlaValLysAsnAlaAsnGluValG 1252
2741 T 2741
1252 u 1252

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:us-08-728-470-10

seq_documentation_block:
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington

STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER TYPE: Floppy disk
MEDIUM TYPE: PC-DOS/MS-DOS
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728.470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302.832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-10

alignment_scores:
Quality: 140.50 Length: 680
Ratio: 0.431 Gaps: 29
Percent Similarity: 47.941 Percent Identity: 20.882

alignment_block:
US-09-049-696-20 x US-08-728-470-10 ..
Align seg 1/1 to: US-08-728-470-10 from: 1 to: 1529

865 GACCTTTAAGAAACCACTCTCTATGACAAACACAGCCACCAATCCACCTT 914
|||::: |||||:::
14 AspGlyAsnLysThrThrIleArgAsnSerValAsnAlaIleIleAsnTr 30
915 CTCATTGCTGCAGATTGGACAAAGAATTTGTGTTTGTAGTCTTGACAAAT 964
|::: |||||:::
30 pLysGlnPheAsnIleAspGlnAsnGluMetGluGlnPheLeuGlnGlu 47
965 CTGCAACCATGGCGACTGTGTAACCGCTCAATCGA..... 999
|::: |||||:::
47 erSerAsnSerAlaValPheAsnArgValThrSerAspGlnIleSerGln 63
1000CTGAATCAAGCAGCCAGCTTTCTCTGCTGCAG..... 1032
64 LeuLysGlyIleLeuAspSerAsnGlyGlnValPheLeuIleAsnProAs 80
1033ACAGTTGAGCTGGGTCCTGGTGGGATGGTGACATTGTCAC. 1074
|||::: |||||:::
80 nGlyIleThrIleGlyLysAspAlaIleIleAsnThrAsnGlyPheThrA 97
1075 ..AGTCTCTCCCATGTACAAAGTGAAGTCACTACAGATAAACAGTGGCAGT 1122
|::: |||||:::
97 laSerThrLeuAspIleSerAsnGluAsnIleLysAlaArgAsnPheThr 113
1123 GACAGGGACACACTCCGCCAAAGATTACCTGACAGCAGCTTCAGGAGGAC 1172
::: |||||:::

```
114 LeuGluGlnThrLysAspLysAlaLeuAlaGluLeuValAsnHisGlyLe 130
1173 GTCCATCTGCAGCGGGTTCGATCGGATTTACTGTGTATAGGAAGAAAT 1222
130 u...ilethrValGlyLysAspGlySerValAsnLeuIleGlyLysV 146
1223 ATCCAACTCAT.....GGATCTGAAATTTGCTGCTG 1254
146 aLysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeuLeu 162
1255 ACGGATGGGAGAACACACTATAAGTGGGTCTTTAACAGGTCACAA 1304
163 Ala...GlyGlnLysIleThrIleSerAspIleIleAsnPro..... 175
1305 AAGTGGTGCATATCCACACAGTCGCTTTGGGGCCCTCTCCAGCTCAAG 1354
176 .....ThrIleThrTyrSerIleAlaAlaProGluAsnGluAlaIleA 190
1355 AACTAGAGAGCTGCCAAATGACAGGAGTTTACAGACATATGCTTCA 1404
190 snLeuGlyAspIlePheAlaLysGlyGlyAsnIleAsnValArgAlaAla 206
1405 GATCAAGTTCAGAACATGGCCCTCAATT.....GATGCTTTTGGGCCCT 1448
207 ...ThrIleArgAsnLysGlyLysLeuSerAlaAspSerValSerLysAs 222
1449 TTCATCAGGAATGAGGTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTA 1498
222 pLysSerGlyAsnIleValLeuSerAlaLysGluGlyGluAlaGluIleG 239
1499 AGGGA...TTAACCCCTCCAGAACACCCAGTGGATGAATGGCAGTGATC 1545
239 lyGlyValIleSerAlaGlnAsnGlnAlaLysGlyGlyLysLeuMet 255
1546 GTG..... 1548
256 ileThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLeuSe 272
1549 .....GACAGCACCGTGG 1561
272 rGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGluArgGlyGluG 289
1562 GAAAGACACTTGTTCATTATCACCCTGCACACAGCAGCCTCCCCAAATC 1611
289 lyLysAsnGlyIleGlnLeuAlaLysLysThrThrLeuGluLysGlySer 305
1612 CTCTCTGGATCCCAAGTGCACAGACAGCAAGGTGGC..... 1647
306 ThrIle...AsnValSerGlyLysGluLysGlyArgAlaIleValTr 321
1648 .....TTGTGTAGTGACAAAACACCAAAATGGCTACCTCCAAA 1687
321 pGlyAspIleAlaLeuIleAspGlyAsnIleAsn.....AlaGlnG 335
1688 TCCAGGCAATGCTAAGTTGGCACATTGGAATACAGTCTGCAAGCAAGC 1737
335 lySerAspIleAlaLysThrGlyGlyPhe.....ValGluThrSer 348
1738 TCACAAACCTTGACCCCTG..... 1755
349 GlyHisAspLeuSerIleGlyAspAspValIleValAspAlaLysGluTr 365
1756 .....ACTGTCACGTCCTCGCTGCCT 1774
365 pLeuLeuAspProAspAspValSerIleGluThrLeuThrSerGlyArgA 382
1775 CCAATGCTACCCCTGCCCTCCCAATACAGTGTACTTCCAAAACGAACAGGAC 1824
382 snAsnThrGlyGluAsnGlnGlyTyrThrThrGlyAspGlyThrLysGlu 398
1825 ACCAGCAAA.....TTCCCCAGCCCTCTGGTAGTTTATGCAATAT 1865
399 SerProLysGlyAsnSerIleSerLysProThrLeuThrAsnSerThrLe 415
```

```
1866 TCGCCAAAGGAGCCTCCCAATTTCTCAGGCCAGTGTCTACAGCCCTGATTG 1915
415 uGluGln.....IleLeuArgArgGlySerTyrValAsnIle 428
1916 AATCACTGAATGGAACAAACAGTTACCTTCGCAACTA...CTGCATTAATCGA 1962
428 hrAlaAsnAsnArgIleTyrValAsnSerIleAsnLeuSerAsnGly 444
1963 GCAGGTGCTGATCTACTAAGGATGACGTGTCTACTCAAGGATATTTCAC 2012
445 SerLeuThrLeuHisThrLysArgAspGlyValLysIleAsnGlyAspIle 461
2013 AACTTATGACACGAATGGTAGATACAGTGTAAAGTCGGGCTCTGGGAG 2062
461 eThrSerAsnGluAsnGlyAsnLeuThrIleLysAlaGlySerTrpValA 478
2063 CAGTTAACGCCAGCCAGCAGGAGATATACCCAGCAGAGTGGAGCACGTG 2112
478 spValHis.....LysAsnIleThrLeuGlyThrGlyPheLeu 490
2113 TACATACCTGGCTGGATTGAGAAATGATGAATACAATGAATCCACCAG 2162
491 AsnIle.....ValAlaGlyAspSerValAlaPheGluArgGlu 504
2163 ACCTGAA..... 2169
504 yAspLysAlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrI 521
2170 .....ATTAATAAGGATGATGTTCAACACACAG.....CAAGTGTGTTC 2208
521 leThrValAsnLysAspLysGlnPheArgPheAsnAsnValSerIle 537
2209 AGCAGAATCCTCGGGAGGCTCATTTGTGGCTTCTGATGTCCCAATGC 2258
538 AsnGlyThrGlyLysGlyLeuLysPheIleAlaAsnGln..... 550
2259 TCCCATACCTCATCTCTTCCACCTGGCCAAATCACCC.....GACC 2299
551 .....AsnAsnPheThrHisLysPheAspG 559
2300 TGAAGCGGAAATTCACGGGGCAGTCTCATTAATCTGACT..... 2340
553 lyGluLeuAsnIleSerGlyIleValThrIleAsnGlnThrThrLysLys 575
2341 .....TGCACAGCTCTCGGGAGTATTGACCAT..... 2370
575 AspValLysTyrTrpAsnAlaSerLysAspSerTyrTrpAsnValSerSe 592
2371 .....GGAACAGCTCACAAAGTAT.....ATCATTCGAA 2398
592 rLeuThrLeuAsnThrValGlnLysPheThrPheIleLysPheValAspS 609
2399 TAAGTACAAAGTATTCTTGATCTCAGACACAGTTCAATGAATCTTTCAA 2448
609 erGlySerAsnSerGlnAspLeuArgSerSerArgArgSerPheAlaGly 625
2449 GTGATACTACTGCTCTCATCCCAAGGAA.....GCCAA 2483
626 ValHisPheAsnGlyIleGlyLysThrAsnPheAsnIleGlyAlaAs 642
2484 CTCTGAGGAAGTCTTTTGTAAACACGAAACAAATTTACT 2523
642 nAlaLysAlaLeuPheLysLeuLysProAsnAlaAlaThr 655
seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.US-08-719-641-10
seq_documentation_block:
: Sequence 10, Application US/08719641
: Patent No. 6218141
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
```

```

; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESS: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-719-641-10

```

```

alignment_scores:
  Quality: 140.50      Length: 680
  Ratio: 0.431        Gaps: 29
  Percent Similarity: 47.941  Percent Identity: 20.882

```

```
alignment_block:

```

```
US-09-049-696-20 x US-08-719-641-10 ..
```

```
Align seg 1/1 to: US-08-719-641-10 from: 1 to: 1529
```

```

865 GACTTTTAAGAAACCACTCTATGACAAACACAGCCACCAATCCCACTT 914
||| |||||
14 ASPLYSnLysThrThrIleArgAsnSerValAsnAlaIleAsnTr 30
||| |||||
915 CTCATGCTGCAGATTGGCAAGAAATTTGTGTTAGTCTTGCACAAAT 964
||| |||||
30 pLysGlnPheAsnIleAspGlnAsnGluMetGluGlnPheLeuGlnGlu 47
||| |||||
965 CTGGAGCATGGGACTGGTAACCGCTCAATCGA..... 999
||| |||||
47 erSerAsnSerAlaValPheAsnArgValThrSerAspGlnIleSerGln 63
||| |||||
1000 .....CTGAATCAAGCCAGCCCTTTCTGCTGCAG..... 1032
||| |||||
64 LeuLysGlyLeuLeuAspSerAsnGlyGlnValPheLeuIleAsnProAs 80
||| |||||
1033 .....ACAGTTGAGCTGGGTCCTGGTGGTGGTGGTGGTGGTGGTGGT 1074
||| |||||
80 nGlyIleThrIleGlyLysAspAlaIleIleAsnThrAsnGlyPheThra 97
||| |||||

```

```

1075 ..AGTCTGCCCCATGTACAAAGTGAAGTACTATACAGATAAACACAGTGGCAGT 1122
||| |||||
97 laSerThrLeuAspIleSerAsnGluAsnIleLysAlaArgAsnPheThr 113
||| |||||
1123 GACAGGAGACACTCGCCCAAAAGATTACCTGCGACGAGCTTCAGAGGGAC 1172
||| |||||
114 LeuGluGlnThrLysAspLysAlaLeuAlaGluIleValAsnHisGlyLe 130
||| |||||
1173 GTCCATCTGCAGCGGCTTCGATCGCATTTACTGTGATTAGGAAGAAAT 1222
||| |||||
130 u...IleThrValGlyLysAspGlySerValAsnLeuIleGlyLysV 146
||| |||||
1223 ATCCAACATGAT.....GGATCTGAATTTGCTGCTG 1254
||| |||||
146 alLysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeuLeu 162
||| |||||
1255 ACGGATGGGGAAGACACACTATAAGTGGTGGTCTTTAACGAGGTCAACA 1304
||| |||||
163 Ala...GlyGlnLysIleThrIleSerAspIleIleAsnPro..... 175
||| |||||
1305 AGTGTGTCATCATCCACACAGCTGCTTTGGGGCCCTCTGCAGCTCAAG 1354
||| |||||
176 .....ThrIleThrTy:SerIleAlaAlaProGluAsnGluAlaIleA 190
||| |||||
1355 AACTAGAGGAGCTGTCCAAATGACAGGAGGTTTACAGACATATGCTTCA 1404
||| |||||
190 snLeuGlyAspIlePheAlaLysGlyGlyAsnIleAsnValArgAlaAla 206
||| |||||
1405 GATCAAGTTCCAGAACAAATGCCTCAT.....GATGCTTTTGGGCCCT 1448
||| |||||
207 ...ThrIleArgAsnLysGlyLysLeuSerAlaAspSerValSerLysAs 222
||| |||||
1449 TTCATCAGGAATGGAGCTGCTCTCAGCGCTCCATCCAGCTTGAGAGTA 1498
||| |||||
222 pLysSerGlyAsnIleValLeuSerAlaLysGlyGlyGluAlaGluIleG 239
||| |||||
1499 AGGA...TTAACCTCCAGACAGCAGTGATGAATGCCACAGTGATC 1545
||| |||||
239 lyGlyValIleSerAlaGlnAsnGlnAlaLysGlyGlyLysLeuMet 255
||| |||||
1546 GTG..... 1548
||| |||||
255 IleThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLeuSe 272
||| |||||
1549 .....GACAGCACCGTGG 1561
||| |||||
272 rGlyLysGluGlyGlyGluThrThrLeuGlyGlyAspGluArgGlyGluG 289
||| |||||
1562 GAAAGGACACTTTGTTTCTTATCACCCTGGACAGCGAGCTCCCAATC 1611
||| |||||
289 lyLysAsnGlyIleGlnLeuAlaLysLysThrThrLeuGluLysGlySer 305
||| |||||
1612 CTTCCTCTGGGATCCCGTGGACAGAGAGGTGGC..... 1647
||| |||||
305 ThrIle...AsnValSerGlyLysGluLysGlyArgAlaIleValTr 321
||| |||||
1648 .....TTGTAGTGACAAACACCAAAATGGCTACCTCCAAA 1687
||| |||||
321 pGlyAspIleAlaLeuAspGlyAsnIleAsn.....AlaGlnG 335
||| |||||
1688 TCCAGGCAATTGTAAGTTGGCATTGGAAATACAGTCTGCAAGCAAGC 1737
||| |||||
335 lySerAspIleAlaLysThrGlyGlyPhe.....ValGluThrSer 348
||| |||||
1738 TCACAAACCTTGACCTG..... 1755
||| |||||
349 GlyHisAspLeuSerIleGlyAspValIleValAspAlaLysGluTr 365
||| |||||
1756 .....ACTGTCACGTCCTCGTGC 1774
||| |||||
365 pLeuLeuAspProAspAspValSerIleGluThrLeuThrSerGlyArgA 382
||| |||||

```

```
1775 CCAATGCTACCTGCTCCATTAACAGTACAGTCTTCCAAAGCAAGGAC 1824
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
382 sAsnThrGlyGluAsnGlnGlyThrThrGlyAspGlyThrLysGlu 398
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1825 ACCAGCAA.....TTCACAGCCCTCTGTAGTTATGCAATAT 1865
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
399 SerProLysGlyAsnSerIleSerLysProThrLeuThrAsnSerThrLe 415
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1866 TGCACAGGAGCTCCCAATTCAGGCGGACAGTGTACAGCCCTGATTG 1915
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
415 uGluGln.....IleLeuArgArgGlySerTyrValAsnIleT 428
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1916 AATCAGTGAATGGAACAGAGTACCTTGGAACTA...CTGATAATGGA 1962
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
428 hrAlaAsnAsnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGly 444
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1963 GCAGTGTGTGCTACTAAGAGTACAGGCTGTCTACTCAAGTATTTCAC 2012
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
445 SerLeuThrLeuHisThrLysArgAspGlyValLysIleAsnGlyAspI 461
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2013 AACTTATGACAGGAATGGTAGATAGTGTAAAGTGGGGCTCTGGGAG 2062
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
461 eThrSerAsnGluAsnGlyAsnLeuThrIleLysAlaGlySerTrpVala 478
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2063 GAGTTAACCCAGCCAGAGGAGAGTATACCCAGCAGAGTGGAGCACTG 2112
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
478 spValHis.....LysAsnIleThrLeuGlyThrGlyPheLeu 490
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2113 TACATACCTGGCTGATTGAGATGATGAATATACATGGAATCCACCAAG 2162
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
491 AsnIle.....ValAlaGlyAspSerValAlaPheGluArgGluG 504
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2163 ACCTGAA..... 2169
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
504 yAspLysAlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrI 521
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2170 .....ATTAATAGGATGATGTTCAACAAAG.....CAAGTGTGTTTC 2208
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
521 leThrValAsnLysAspAspLysGlnPheArgPheAsnAsnValSerIle 537
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2209 AGCAGAACATCTCGGAGGCTCATTTGTGGCTTCTGATGTCCCAATGC 2258
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
538 AsnGlyThrGlyLysGlyLeuLysPheIleAlaAsnGln..... 550
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2259 TCCCATACCTGATCTCTCCACCTGGCGCAATCACC.....GACC 2299
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
551 .....AsnAsnPheThrHisLysPheAspG 559
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2300 TGAAGCGGAAATTACGGGGCAGCTCTATTAACTGACT..... 2340
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
559 yGluLeuAsnIleSerGlyIleValThrIleAsnGlnThrThrLysLys 575
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2341 .....TGGACAGCTCTGGGGATGATTATGACCAT..... 2370
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
576 AspValLysTyrTrpAsnAlaSerLysAspSerTyrTrpAsnValSerSe 592
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2371 .....GGACAGCTCACAGTAT.....ATCATTCGAA 2398
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
592 rLeuThrLeuAsnThrValGlnLysPheThrPheIleLysPheValAspS 609
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2399 TAAGTACAAGTATCTGTATCTCAGAGACAAGTTCAATGAATCTCTCAA 2448
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
609 erGlySerAsnSerGlnAspLeuArgSerSerArgArgSerPheAlaGly 625
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2449 GTGAATACTACTGCTCTCATCCCAAGGAA.....GCCAA 2483
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
626 ValHisPheAsnGlyIleGlyLysThrAsnPheAsnIleGlyAlaAs 642
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2484 CTCTGAGGAAGTCTTTTGTGTTTAAACACAGAAACATTACT 2523
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
642 nAlaLysAlaLeuPheLysLeuLysProAsnAlaAlaThr 655
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.us-08-038-682-2

```
seq_documentation_block:
; Sequence 2, Application US/08038682
; Patent No. 5549897
;
GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
;
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
;
INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-038-682-2
```

```
alignment_scores:
  Quality: 138.50      Length: 936
  Ratio: 0.318        Gaps: 49
  Percent Similarity: 46.581      Percent Identity: 20.406
```

alignment_block:

US-09-049-696-20 x US-08-038-682-2 ..

Align seg 1/1 to: US-08-038-682-2 from: 1 to: 1536

```
217 TTTGAAGCTACA.....GGAAGCGATTTTATTTCAAAAATGT 254
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnIle 643
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
255 TGCCATTTTGATTCTCGTGAACATGGAACAAAGGCTGACTATGTGAGAC 304
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
643 eSerMetValLeuPro.....LysAsnGluSerGlyTyrAspLysP 657
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
305 CAAAACTTGAGACCTACAAAAATGCTGATGTTCTG...GTTGCTGAGTCT 351
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
657 heLysGlyArgThrTyrTrpAsnLeuThrSerLeuAsnValSerGluSer 673
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
352 .....ACTCCTCCAGTAAATGAT..... 369
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
674 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspAlaGlyTh 690
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
370 .....GAACCCCTACACT.....GAGCAGA 388
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
690 rLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThrT 707
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```


1108ATAACAGTGGCAGTGACACGGGACACACTCGCCCAAGA 114
992 ysGluclyAsnLeuThrIleSerSerIspIlySIIeAasnlleThrLysGln 1008
1147TTACCTGCAGCAGCTTCAGGAGGGAGCTCCATCTGCAGCGGGCT 1190
1005 IleThrIleLysAlaglyValaspGlyGluasnSerAspserAspAlaThr 1025
1191 TCGATCGGCATTACTGTGATTAGGAAGA... 1221
1025 rAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnA 1042
1222TATCCACTGATGGA 1236
1042 spLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysASPgly 1058
1237 TCCTGAATTTGCTCTGCTCACGGATGGGACACACACTATAAGTGG... 1284
1059 SerAsp.....LeuThrIleGlyAsnThrAsnSerAlaaspGlyTh 1072
1285TCCTTTAACGAGGTCAAACAAGT.....GGTG 1312
1072 rAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089
1313 CCATCATCCACACAGCTCGCTTTGGGCCCTCTGCAGCTCAAGAACTAGAG 1362
1089 IaaspGlyHisLysValThrLeuHis..... 1097
1363 GAGCTGTCCAAAATG...ACAGGAGTGTTCACAGACATATGCTTCAGATCA 1409
1098SerLysValGluThrSerGlySerAsnAsnAsnThrGluaspse 1112
1410 AGTTCAACAATGCGCTCATGTGCTTTTGGGGCCCTTTCATCAGGAA 1459
1112 rSerAspAsnAsnAla..... 1117
1460 ATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAGGAGTAACC 1509
1118GlyLeuThrIleAspAlaLysAsnValThr 1127
1510 CTCAGAACAGCAGCTGGATGAATGGCACAGCTGATCTGGACAGCACCGT 1559
1128 ValasnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144
1560 GGAAAGGACACATTTGTTTCTTATCACCTGCAGCAACAGCCCTCCCCAAA 1609
1144 rGlyGlu.....IleThrThrLysThr..... 1151
1610 TCCTTCTCTGGGATCCCAGTGGACAGAAGAGTGCTGTGTAGTGGAC 1659
1152GlyThrThrIleAsn 1156
1660 AAAACACCAAAATGGCTACCTCCAATCCAGGCATTCCTAAGGTGG 1709
1157 AlaThrThrGly.....AsnValGluIleThrAlaGlnThrGl 1169
1710 CACTTGGAAATACAGTCTGCAAGCAAGCTACAAACCTTCACCTGACTG 1759
1169 ySerIleLeuGlyGlyIleGluSerSerSerGlySerValThrLeuThrA 1186
1760 TCAGTCCGGTGGC.....TCCAATGTACCTCGCTCCCAATTACA 1800
1186 laThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
1801 GTGACTTCCAAACCAACAGGACACAGC..... 1830
1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaglySerThrIleLy 1219
1830 1830
1219 sGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyClyTr 1236
1831AAATTCACGACCCCTCTGTAGTT 1854

```
1236 hrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThr 1252
      |||      :|||:
1855 TATGCAATATTCG.....CAAGAGCCTCCCAATCTCAG 1892
      :|||: :|||:
1253 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 1269
      :|||: :|||:
1893 GGCCAGTCTCAGCCCGGATTCAGTCAATGAGTGAAGAAACAGTTCACCT 1942
      :|||: :|||:
1269 rAlaThrGlyThr.....IleGlyGlyThrIleSerGlyAsnThrValAsnV 1285
      :|||: :|||:
1943 TGGAA.....CTACTGGATAATGGACGAGGTGCTCAT 1974
      :|||: :|||:
1285 alThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 1301
      :|||: :|||:
1975 GCTACTAG.....GATGACGGTGTCTACTCAAG 2003
      :|||: :|||:
1302 AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 1318
      :|||: :|||:
2004 GTATTTCAACATTATGACAGCAATGGTGTAGTACAGTGTAAAGTGGCGG 2053
      :|||: :|||:
1318 rGluAlaSerSerHisIleThrSerAlaLysGlyGlnValAsnLeuSerA 1335
      :|||: :|||:
2054 CT.....CTGGGAGGAGTTAACGCCAGCCAGCAGGAGGTG 2088
      :|||: :|||:
1335 laGlnAspGlySerValAlaGlySerIleAsnAlaAla.....AsnVal 1349
      :|||: :|||:
2089 ATACCCACGAGGAGTGGAGCACTGTACATCTGCTGGTGGATGAGAAATGA 2138
      :|||: :|||:
1350 ThrLeuAsnThrThrGlyThrLeuThrThr.....ValLysGlySe 1363
      :|||: :|||:
2139 TGAATATCAATGAATCCACCAAGACCTGAAATTAATGAAGATGATGTC 2188
      :|||: :|||:
1363 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 1380
      :|||: :|||:
2189 AA.....CACAGCAAGTGTGTTTCAGCAGA 2214
      :|||: :|||:
1380 lueuAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn 1396
      :|||: :|||:
2215 ACATCCTCGGAGGCTCATTTGTGGCTTCTGTATGTCCTCCAAATGCTCCCAT 2264
      :|||: :|||:
1397 AlaAsnGlySerGlySerValIleAlaThrThrSerSerArgValAsnIl 1413
      :|||: :|||:
2265 ACCT...GATCTCTCCCACTGGCCCAATC.....ACCG 2296
      :|||: :|||:
1413 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleIleSerLysAsnG 1430
      :|||: :|||:
2297 ACCTGAAGCGGAAATTCACGGGGCAGTCTCATTAATCTGACTTGGACA 2346
      :|||: :|||:
1430 lYleAsnThrValLeuLeuLysGlyValLysIleAspValLysTyrlle 1446
      :|||: :|||:
2347 GCTCCTGGG...GATGATTATGACCATGGAACAGCTCACAGTATATCAT 2393
      :|||: :|||:
1447 GlnProGlyIleAlaSerValAspGluValIleGluAlaLys..... 1460
      :|||: :|||:
2394 TCGAATAAGTACAGATTCCTGATCTCAGACAGCAAGTTCATGATATCTC 2443
      :|||: :|||:
1461 ArgIleLeuGluLysValLysAspLeuSerAspGluGluArgGluAlaL 1477
      :|||: :|||:
2444 TT 2445
      :|||:
1477 eu 1477

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-302-832-2
seq_documentation_block:
; Sequence 2, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; DATE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
```

```
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302.832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION NUMBER: US pct/us93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

alignment_scores:
Quality: 138.50 Length: 936
Ratio: 0.318 Gaps: 49
Percent Similarity: 46.581 Percent Identity: 20.406

alignment_block:
US-09-049-696-20 x US-08-302-832-2
Align seg 1/1 to: US-08-302-832-2 from: 1 to: 1536
217 TTTCAGCTACA.....GGAAGCGATTTTATTTCAAAAATGT 254
      |||||:|||||:
631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnIl 643
255 TCCATTTTGTATTCCTGAAACATGGAACAAAGCTGACTATGTGAGAC 304
      |||||:|||||:
643 eSerMetValLeuPro.....LysAsnGluSerGlyTyrrAspLysP 657
305 CAAACITGAGACCTACAAAATCTGTGTTCTG...GTGCTGAGTCT 351
      |||||:|||||:
657 heLysGlyArgThrTyrrpAsnLeuThrSerLeuAsnValSerGluSer 673
352 .....ACTCCTCCAGGTAATGAT..... 369
674 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspSerAlaGlyTh 690
370 .....GAACCCCTACACT.....GAGCAGA 388
      :|||: :|||:
690 rLeuThrGlnProTyrrAsnLeuAsnGlyIleSerPheAsnLysAspThrT 707
369 TGGGCAACTGTGGAGAGA.....AGGTGAAAGGATCCACCTCCTCCT 432
      ||| |||||: |||||:
707 hrPheAsn.ValGluArgAsnAlaArgValAsnPheAspIleLysAlaPr 723
```



```

1 STREET: Bldg. 1
2 CITY: Arlington
3 STATE: Virginia
4 COUNTRY: U.S.A.
5 ZIP: 22202-0286
6 COMPUTER READABLE FORM:
7 MEDIUM TYPE: Floppy disk
8 COMPUTER: IBM PC compatible
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10 SOFTWARE: PatentIn Release #1.0, Version #1.25
11 CURRENT APPLICATION DATA:
12 FILING DATE: 13-DEC-1995
13 CLASSIFICATION: 424
14 ATTORNEY/AGENT INFORMATION:
15 NAME: BERKSTRESSER, JERRY W
16 REGISTRATION NUMBER: 22,651
17 REFERENCE/DOCKET NUMBER: JWB-1186
18 TELECOMMUNICATION INFORMATION:
19 TELEPHONE: (703) 415-0810
20 TELEFAX: (703) 415-0813
21 INFORMATION FOR SEQ ID NO: 2:
22 SEQUENCE CHARACTERISTICS:
23 LENGTH: 1536 amino acids
24 TYPE: amino acid
25 STRANDEDNESS: single
26 TOPOLOGY: linear
27 MOLECULE TYPE: protein
28 US-08-530-198-2
29
30 alignment_scores:
31 Quality: 138.50 Length: 936
32 Ratio: 0.318 Gaps: 49
33 Percent Similarity: 46.581 Percent Identity: 20.406

```

```

alignment_scores:
  Quality: 138.50      Length: 936
  Ratio: 0.318        Gaps: 49
  Percent Similarity: 46.581      Percent Identity: 20.406

alignment_block:
  US-09-049-696-20 x US-08-530-198-2 ..
Align seq 1/1 to: US-08-530-198-2 from: 1 to: 1536

```

217 TTGAAAGCTACA.....GGAAGCGATTATTATTTCAAAATGT 254
 |||||:|||||
 631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnI 643
 255 TGCATTTTGTATTCCTGAAACATGGAAGACAAAGGCTCACTATGTGAGAC 304
 :|||:|||||
 643 eSerMetValLeuPro.....LysAsnGluSerGlyTyrAspLysP 657
 305 CAAAACTTGAGACCTACAAAATGCTGATGTTCTG...GTTGCTGAGTCT 351
 |||:|||||
 657 helGlyArgThrTyrTrrPAsnLeuThrSerLeuAsnValSerGluSer 673
 352.....ACTCCTCCAGGTAATGAT..... 369
 |||:|||||
 674 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspSeraGlyTh 690
 370.....GAACCTTACACT.....GACCAGA 388
 :|||||:|||||
 690_rLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThr 707
 389 TGGGCAACTGTGGAGAGA.....AGGGTGAAGGATCCACCTCACTCCT 432
 :|||:|||||
 707_hrPheAsn_ValGluArgAsnAlaArgValAsnPheAspIleLysAlaPr 723
 433 GATTTTCATTCAGGMAAAAGTTAGCTGAATATGGACCACACAGGTAGGC 482
 |||||:|||||
 723_olLeGlyIleAsnLysTyrSerSerLeuAsnTyrAlaSerPheAsnGlyA 740
 483 ATTTGTTCATGAGTGGGCTCATCTACGATGGGAGCTATTTGACGAGTACA 532
 :|||||:|||||
 740_snIleSerValSerGly.....GlyGlySerValAspPheThr 752

```

seq_documentation_block:
: Sequence 2, Application US/085330198
: Patent No. 5869065
: GENERAL INFORMATION:
: APPLICANT: BARENKAMP, STEPHEN J
: APPLICANT: ST. GEME III, JOSEPH W
: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
: TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

```

533 ATAATGATGAGAAATCTTACTTATCCAAATGGAACAATACA.....AGCA 576
573 LeuLeuAlaSerSerAsn.ValGlnThrProGlyValValIleAsnS 769
577 GTAAGATGTTTCAGCAGCAGTATTACTGTGACAAATCTAGTAAGAAGTGTCA 626
579 erLysTyrPheAsn.ValSerThrGlySerSerLeuArgPheLysThrSe 785
627 GGA.....GGCAGCTGTTACACCAAAAGATGCATTCAT. 663
785 rGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeuAsnA 802
664AAGTAACAGGACTCTATGAAAAAGGATGTGAGTTTGT 702
802 laThrGlyLysAsnIleThrLeuLeuGlnValGluGlyThrAspGlyMet 818
703 CTCCAATCCCGCCAGAGAGAGCTTCTATATATGTTT..... 741
819 IleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyGlyAs 835
742GCACACATGTTGATTCATCTATAGTTG 766
835 nIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsnValT 852
767 AATTCCTGTACAGAA.....CAAAAC 786
852 hrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPheAspAsn 868
787 CACAAACAAGAGCTCCAAACAAGCAAAAT.....CAAAATGCAA 827
869 HisGlnLysProLeuThrIleLysLysAspValIleIleAsnSerGlyAs 885
828 TCCTCGAAGCACATGGGAAGTG.....ATCC 853
885 nLeuThrAlaGlyLysIleValAlaAsnIleAlaGlyAsnLeuThrValG 902
854 GTGATTCGTGAGGACTTTAAAGAAACCACTCTCTATGACA..... 891
902 luSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsnValGly 918
892ACACAGCCACCAATCCACCTTCTCATTTGCTGCAGATG 932
919 GlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLysGlyG 935
933 ACAAGAATTTGTGTTAGTCTTGTGACAAATCTGGAAGCATGCCGACTG 982
935 yAlaArgPheLysAspIleAspAsnSerLysAsnLeuSerIleThrA 952
983 GTAACCGCCTC.....AATCGACTGAATCAA 1008
952 snSerSerThrTyrArgThrIleIleSerGlyAsnIleThrAsnLys 968
1009 GCAGGCCAGCTTTTCCTGTCGACAGAGTTGAGCTGGGGTCTCGGTGG 1058
969 AsnGlyAspLeuAsnIleThrAsnGlu.....GlySer..... 979
1059 GATGGTGACATTTGACATGCTGCCCATGTACAAAGTGAAGTCAATACAG. 1107
980AspThrGluMetGlnIleGlyGlyAspValSerGlnL 992
1108ATAACAGTGGGAGTGCACAGGACACACTCGCCAAAGA 1146
992 ysGluGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrLysGln 1008
1147TTAGCTGCAGCAGCTTCAGGAGGACGTCCATCTGCAGCGGGCT 1190
1009 IleThrIleLysAlaGlyValAspGlyGluAsnSerAspAspAlaThr 1025
1191 TCGATCGGCATTTACTGTGATTTAGGAAGAA..... 1221
1025 rAsnAsnAlaAsnLeuThrIleThrLysThrLysGluLeuLysLeuThrGlnA 1042
1222TATCCAACTGATGGA 1236

1042 spLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGly 1058
1237 TCTCAAAATTTGCTGCTGCGGATGGGAAGACAACTATAAGTGGG... 1284
1059 SerAsp.....LeuThrIleGlyAsnThrAsnSerAlaAspGlyTh 1072
1285TGCTTTAACGAGGTCAACAAGT.....GGTG 1312
1072 rAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089
1313 CCATCATCCACACAGTCGCTTTGGGCCCTCTGCAGCTCAAGAACTAGAG 1362
1089 laAspGlyHisLysValThrLeuHis..... 1097
1363 GAGCTGTCCAAATG...ACAGGAGGTTTACAGACATATGCTTCAGATCA 1409
1098SerLysValGluThrSerGlySerAsnAsnThrGluAspSe 1112
1410 AGTTCAGAACAAATGGCTCATGCTTTTGGGCCCTTTTCATCAGGAA 1459
1112 rSerAspAsnAla..... 1117
1460 ATGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAAC 1509
1118GlyLeuThrIleAspAlaLysAsnValThr 1127
1510 CTCAGAACAGCCAGTGGATGAATGGCAGATGATCGTGGACAGACCCGT 1559
1128 ValAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144
1560 GGAAGGACACTTTGCTTATACCTGGACAACCCACCTCCCAAA 1609
1144 rGlyGlu.....IleThrThrLysThr..... 1151
1610 TCCTTCTCTGGGATCCAGTGGACAGAAGCAAGTGGCTTTGTAGTGGAC 1659
1152GlyThrThrIleAsn 1156
1660 AAAACACCAAAATGGCTTACCTCCAAATCCAGGCATTGCTAAGTTGG 1709
1157 AlaThrThrGly.....AsnValGluIleThrAlaGlnThrG 1169
1710 CACTTGGAAATACAGTCTGCAACAGCTCACAACCTTGACCTGACTG 1759
1169 ySerIleLeuGlyGlyIleGluSerSerGlySerValThrLeuThrA 1186
1760 TCAGTCCCGCTCG.....TCCAATGCTACCTGCTCCCAATTACA 1800
1186 laThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
1801 GTGACTTCCAAACGAAACAGGACACCAGC..... 1830
1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLy 1219
1830 1830
1219 sglyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyGlyT 1236
1831AAATTCGCCCTCTCGTAGTT 1854
1236 hrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThrThr 1252
1855 TATGCAATATTCGC.....CAAGGAGCCTCCCAATTCCTCAG 1892
1253 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 1269
1893 GGCAGTGTCCAGCCCTGATTAATCAGTGAATGGAACACAGTTACCT 1942
1269 rAlaThrGlyThr...IleGlyGlyThrIleSerGlyAsnThrValAsnV 1285
1943 TGGAA.....CTACTGGATTAATGAGCAGGCTGTGAT 1974
1944 1974


```

753 LeuLeuAlaSerSerSerAsn.ValGlnThrProGlyValValIleAsnS 769
      ::::: ::::: ::::: ::::: :::::
577 GTAAGATGTCAGCAGGTATTACTGGTACAAATGATGTAAGAAGTGCA 626
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
769 erLysTyrPheAsn.ValSerThrGlySerSerLeuArgPheLysThrSe 785
      ::::: ::::: ::::: ::::: :::::
627 GGA .....GCACCTGTTACACCAAAAGATGCACATCAAT. 563
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
785 rGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeuAsnA 802
      ::::: ::::: ::::: ::::: :::::
664 .....AAACTAACAGCAGCTCTATGAAAGAGGATGTCAGTTGTT 702
      ::::: ::::: ||||| ||||| ||||| ||||| |||||
802 laThrGlyGlyAsnIleThrLeuGlnValGluGlyThrAspGlyMet 818
      ::::: ::::: ||||| ||||| ||||| ||||| |||||
703 CTCCAATCCCGCAGCAGGAGGAGCTTCTATAATGTTT ..... 741
      ::::: ::::: ||||| ||||| ||||| ||||| |||||
819 ileGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyGlyAs 835
      ::::: :::::
742 .....GCACAACATGTTGATTCATATAGTTG 766
      ::::: ::::: ||||| ||||| ||||| ||||| |||||
835 nIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsnValT 852
      ::::: :::::
767 AATTCGTACAGAA .....CAAAC 786
      ::::: :::::
852 hrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPheAspAsn 868
      ::::: :::::
787 CACACAAGAAGCTCCAAACAAAGCAAAAT .....CAAATGCAA 827.
      ||||| ||||| ||||| ||||| ||||| |||||
869 HisGlnLysProLeuThrIleLysLysAspValIleIleAsnSerGlyAs 885
      ::::: :::::
828 TCTCGAAGCACATCGGAAGTG .....ATCC 853
      ||||| ||||| ||||| ||||| ||||| |||||
885 nLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuThrValG 902
      ::::: ::::: ||||| ||||| ||||| ||||| |||||
854 GTGATTCTGAGGACTTAAAGAAACCTCTCTATGACA ..... 891
      ::::: ::::: ||||| ||||| ||||| ||||| |||||
902 luSerAsnAlaAsnPhelysAlaIleThrAsnPherThrPheAsnValGly 918
      ::::: ::::: ||||| ||||| ||||| ||||| |||||
892 .....ACACGCCACCAATCCACCTCTCTATTGCTGCAGATTGG 932
      ::::: ::::: ||||| ||||| ||||| ||||| |||||
919 GlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLysGlyG1 935
      ::::: :::::
933 ACAAGAATTGTTGTTAGTCTTACAAATCTGGAAGCATGGCGACTG 982
      ||||| ||||| ||||| ||||| ||||| |||||
935 yalaargPhelysAspIleAspAsnSerLysAsnLeuSerIleThrThRA 952
      ::::: :::::
983 GTAACCGCCTC .....AATCGACTGAATCAA 1008
      ::::: ||||| ||||| |||||
952 snSerSerThrTyrArgThrIleIleSerGlyAsnIleThrAsnLys 968
      ::::: ::::: ||||| ||||| ||||| ||||| |||||
1009 GCAGCCAGCTTTCTCTGTCGACAGATTGAGCTGGGGTCTCGGTGG 1058
      ||||| ||||| ||||| ||||| ||||| |||||
969 AsnGlyAspLeuAsnIleThrAsnGlu .....GlySer. 979
      ::::: :::::
1059 GATGGTGACATTTGACAGTGTGCCCTACAAAGTCAACTCATACAG. 1107
      ||||| ||||| ||||| ||||| ||||| |||||
980 .....AspThrGluMetGlnIleGlyAspValSerGlnL 992
      ::::: :::::
1108 .....ATAACAGTGGCAGTGACAGGCACACTCGCCAAAAGA 1146
      ::::: ::::: ||||| ||||| ||||| ||||| |||||
992 ysGluGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrLysGln 1008
      ::::: ::::: ||||| ||||| ||||| ||||| |||||
1147 .....TTACCTGCAGCAGCTTCAGAGGAGCGTCCATCTCGACGGGCT 1190
      ::::: ||||| ||||| ||||| ||||| |||||
1009 lleThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAlaTh 1025
      ::::: :::::
1191 TCGATCGGCATTTACTGTCATTAGGAAGAAA ..... 1221
      ::::: ||||| ||||| ||||| ||||| |||||
1025 rAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnA 1042
      ::::: :::::
1222 .....TATCCAACTCATGGA 1236
      ||||| |||||

```

```

1042 spLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGly 1058
      ::::: :::::
127 TCTGAAATGTCTGCTGACGATGGGGAAGACAACTATAAGTGGG. 1284
      ||||| ||||| ||||| ||||| ||||| |||||
1059 SerAsp. ....LeuThrIleGlyAsnThrAsnSerAlaAspGlyTh 1072
      ::::: :::::
1285 .....TGCTTTAACGAGGTCAACAAAGT .....GGTG 1312
      ||||| ||||| ||||| ||||| ||||| |||||
1072 rAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089
      ::::: :::::
1313 CCATCATCCACACAGTCGCTTTGGGGCCCTCTGTCAGCTCAAGAACTAGAG 1362
      ||||| ||||| ||||| ||||| ||||| |||||
1089 laAspGlyHisLysValThrLeuHis ..... 1097
      ::::: :::::
1363 GAGCTGTCCAAATG...ACAGGAGGTTTACACACATATGCTTTCAGATCA 1409
      ||||| ||||| ||||| ||||| ||||| |||||
1098 .....SerLysValGluThrSerGlySerAsnAsnThrGluAspSe 1112
      ::::: :::::
1410 AGTTTCAGAACAAATGCCCTCATTTGATGCTTTTGGGGCCCTTTCATCAGGAA 1459
      ::::: ||||| ||||| ||||| ||||| |||||
1112 rSerAspAsnAsnAla ..... 1117
      ::::: :::::
1460 ATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAGGATTAAAC 1509
      ::::: ::::: ||||| ||||| ||||| ||||| |||||
1118 .....GlyLeuThrIleAspAlaLysAsnValThr 1127
      ::::: :::::
1510 CTCCAGAACAGCAGTGGATGAATGGCAGAGTATCGTGGACAGCACCGT 1559
      ::::: ||||| ||||| ||||| ||||| |||||
1128 ValAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144
      ::::: :::::
1560 GGAAGAGCACCTTTGTTCTTATCACCCTGGACACAGCGCTCCCAAA 1609
      ||||| ||||| ||||| ||||| |||||
1144 rGlyGlu. ....IleThrThrLysThr ..... 1151
      ::::: :::::
1610 TCCTTCTCTGGATCCAGTGGACAGAACAGGCTTGTGTAGTGGAC 1659
      ::::: ||||| ||||| ||||| ||||| |||||
1152 .....GlyThrThrIleAsn 1156
      ::::: :::::
1660 AAAACACCAAAATGGCTTACCTCCAAATCCAGGCAATGCTAAAGTTGG 1709
      ::::: ||||| ||||| ||||| ||||| |||||
1157 AlaThrThrGly. ....AsnValGluIleThrAlaGlnThrG1 1169
      ::::: :::::
1710 CACTTGGAAATACAGTCTCAAGCAGAGCTCACAACTTGACCCCTGACTG 1759
      ||||| ||||| ||||| ||||| ||||| |||||
1169 ySerIleLeuGlyGlyIleGluSerSerGlySerValThrLeuThRA 1186
      ::::: :::::
1760 TCAGCTCCCGTCGG .....TCAATGCTACCTGCTCCCAATTACA 1800
      ::::: ||||| ||||| ||||| ||||| |||||
1146 laThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
      ::::: ||||| ||||| ||||| ||||| |||||
1801 GTGACTTCCAAACGAACAGGACACCAAGC ..... 1830
      ||||| ||||| ||||| ||||| ||||| |||||
1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLy 1219
      ::::: ||||| ||||| ||||| ||||| |||||
1830 ..... 1830
      ::::: :::::
1219 sglyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyGlyT 1236
      ::::: :::::
1831 .....AAATCCCGCCAGCCCTCTGGTAGTT 1854
      ||||| ||||| ||||| ||||| ||||| |||||
1236 hrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThrThr 1252
      ::::: ||||| ||||| ||||| ||||| |||||
1855 TATGCAATATTTCG. ....CAAGGAGCTCCCAATTCCTCAG 1892
      ::::: ||||| ||||| ||||| ||||| |||||
1253 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 1269
      ::::: ||||| ||||| ||||| ||||| |||||
1893 GCCAGTGTCAAGCCCTGATTCAATCAGTGAATGAAAAACAGTTACCT 1942
      ||||| ||||| ||||| ||||| ||||| |||||
1269 rAlaThrGlyThr...IleGlyGlyThrIleSerGlyAsnThrValAsnV 1285
      ::::: ||||| ||||| ||||| ||||| |||||
1943 TGGAA .....CTAGTGAATATGAGCAGCGTGTGAT 1974
      ::::: ||||| ||||| ||||| ||||| |||||
1285 aThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 1301
      ::::: ||||| ||||| ||||| ||||| |||||

```

```
1975 GCTACTAAG.....GATGACGGTGTCTTACTACTAAG 2003
|||||:
1302 AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 1318
|||||:
2004 GTATTTCACAACTTATGACACAGAAATGGTAGATACATGTGTAAGAGTCGGG 2053
|||||:
1318 rGluAlaSerSerHisLeThrSerAlaLysGlyGlnValAsnLeuSerA 1335
|||||:
2054 CT.....CTGGGAGGAGTTAAACGACGCCAGACGAGAGTG 2088
|||
1335 laGlnAspGlySerValAlaGlySerIleAsnAlaAla.....AsnVal 1349
|||||:
2089 ATACCCAGCAGAGTGGAGACACTGTACATACCTGGCTGGATTGAGATGA 2138
|||||:
1350 ThrLeuAsnThrThrGlyThrLeuThr.....ValLysGly 1363
|||||:
2139 TGAATACATGAATGCCACCAAGACCTGAAATTAATAAGGATGATGTC 2188
|||||:
1363 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 1380
|||||:
2189 AA.....CACAGCAAGTGTGTTTCAGCAGA 2214
|||
1380 luLeuAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn 1396
|||||:
2215 ACATCTCGGAGGCTCATTTGTGGCTCTCATGTCCCAAATGCTCCCAT 2264
|||||:
1397 AlaAsnGlySerGlySerValIleAlaThrThrSerSerArgValAsnII 1413
|||||:
2265 ACCT...GATCTCTCCACCTGGCCAAATC.....ACCG 2296
|||||:
1413 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleIleSerLysAsnG 1430
|||||:
2297 ACCTGAAGCGGAAATCAGCGGGGCGAGCTCATTAATCTACTGAGTGGACA 2346
|||||:
1430 lyIleAsnThrValLeuLysGlyValLysIleAspValLysIle 1446
|||||:
2347 GCTCTCTGGG...GATGATTATGACCATCGAACAGCTCACAACTATATCAT 2393
|||||:
1447 GlnProGlyIleAlaSerValAspGluValIleGluAlaLys..... 1460
|||||:
2394 TCGAATAAGTACAGTATTCTTGTATCTCAGACACAACTTCATGATCTC 2443
|||||:
1461 ArgIleLeuGluLysValLysAspLysSerAspGluArgGluAlaL 1477
|||||:
2444 TT 2445
1477 eu 1477

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-728-470-2

seq_documentation_block:
; Sequence 2, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```


753 LeuLeuAlaSerSerSerAsn.ValGlnThrProGlyValValIleAsnS 769
 577 GTAAGATGTTCCAGCAGGTATTACTAGTGTACAAATGATGAGTAAGAAGTGCA 626
 769 erLysTyrPheAsn.ValSerThrGlySerSerLeuArgPheLysThrSe 785
 627 GGA.....GGCAGCTGTACACCAAAAGATGCACATCAAT. 663
 785 rGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeuAsnA 802
 664AAGTAACAGGACTCTATGAAAAAGGATGTGAGTTGTT 702
 802 laThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspGlyMet 818
 703 CTCCAATCCGCCAGCAGGAGGAGCTTCTATAATGTTT..... 741
 819 IleGlyLysGlyIleValAlaLysAsnIleThrPheGluGlyGlyAs 835
 742GCACAACATGTTGATTCTATAGTTG 766
 835 nIleThrPheGlySerArgLysAlaValThrGluIleGlyAsnValT 852
 767 AATTCGTACAGAA.....CAAAAC 786
 852 hrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPheAspAsn 868
 787 CACAACAAGAGCTCCAAACAGCAAAAT.....CAAAATGCAA 827
 869 HisGlnLysProLeuThrIleLysLysAspValIleIleAsnSerGlyAs 885
 828 TCTCCGAGCAGCATGGGAGTG.....ATCC 853
 885 nLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuThrValG 902
 854 GTGATTCGTAGGACTTTAAGAAACCACTCTCTATGACA..... 891
 902 luSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsnValGly 918
 892ACACAGCCACCAATCCACCTCTCATTCGTGCGAGATGG 932
 919 GlyLeuPheAspAsnLysGlyAsnSerAsnIleAlaLysGlyG 935
 933 ACAAGAAATGTTGTTAGTCTTGACAAATCTGGAAGCATGGGACTG 982
 935 yAlaArgPheLysAspIleAspAsnSerLysAsnLeuSerIleThrThra 952
 983 GTAACCGCCTC.....AATCGACTGAATCAA 1008
 952 snSerSerThrTyrArgThrIleIleSerGlyAsnIleThrAsnLys 968
 1009 CGAGCCAGCTTTCTGCTGCTGACAGACTTGAGTGGGCTCGGTGG 1058
 969 AsnGlyAspLeuAsnIleThrAsnGlu.....GlySer..... 979
 1059 GATGGTGACATTTGACAGTGTGCCATGTACAAAGTGAACATACAG. 1107
 980AspThrGluMetGlnIleGlyGlyAspValSerGlnL 992
 1108ATAAACAGTGCAGTGACAGGACACACTGCCCAAAAGA 1146
 992 ysGluGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrLysGln 1008
 1147TTACCTGCAGCAGCTTACGAGGAGCGTCCATCTGCAGCGGCT 1190
 1009 IleThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAlaTh 1025
 1191 TCGATCGCATTTACTGTGATTAGGAAGAAA..... 1221
 1025 rAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnA 1042
 1222TATCCAACCTGATGGA 1236
 1042 spLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGly 1058

1237 TCTGAATTTGCTGCTCTGACGGATGGGAAGACAAACACTATAAGTGGG... 1284
 1059 SerAsp.....LeuThrIleGlyAsnThrAsnSerAlaAspGlyTh 1072
 1285TGCTTTAACGAGGTCAACAAAGT.....GGTG 1312
 1072 rAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089
 1313 CCATCATGCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAACAACATAGAG 1362
 1089 laAspGlyHisLysValThrLeuHis..... 1097
 1363 GAGCTGTCCAAATG...ACAGGAGTTTACAGACATATGCTTCAGATCA 1409
 1098SerLysValGluThrSerGlySerAsnAsnThrGluAspSe 1112
 1410 AGTTCAGAACATAGCCCTCATTTGCTTTGGGCCCTTTTCATCAGAA 1459
 1112 rSerAspAsnAsnAla..... 1117
 1460 ATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAOC 1509
 1118GlyLeuThrIleAspAlaLysAsnValThr 1127
 1510 CTCAGAACAGCAGTGGATGAATGGCAGCATGCTGGGAGCAGCACCT 1559
 1128 ValAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144
 1560 GGAAGGAGCACCTTTCTTATCACCCTGGACAACGCGACCTCCCAAA 1609
 1144 rGlyGlu.....IleThrThrLysThr..... 1151
 1610 TCCTTCTCTGGGATCCAGTGGACAGAACAGGTCGCTTTTGTAGTGGAC 1659
 1152GlyThrThrIleAsn 1156
 1660 AAAACACCAAAATGGCTACTCTCCAAATCCAGGCAATGCTAAGTTGG 1709
 1157 AlaThrThrGly.....AsnValGluIleThrAlaGlnThrG 1169
 1710 CACTTGGAAATACAGTCTGCAAGCAAGCTGCACAACTTGACCTGACTG 1759
 1169 ySerIleLeuGlyGlyIleGluSerSerGlySerValThrLeuThrA 1186
 1760 TCAGCTCCCGTGGC.....TCCAATGCTACCTGCTCCCAATTACA 1800
 1168 laThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
 1801 GTGACTTCCAAACACAGACACACACG..... 1830
 1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIle 1219
 1830 1830
 1219 sGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyGlyT 1236
 1631AAATCCCGAGCCCTCTGGTAGTT 1854
 1236 hrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThrThr 1252
 1855 TATGCAAAATTCGC.....CAGGAGCTTCCCAATTCTCAG 1892
 1253 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 1269
 1893 GSCCAGTGTACAGCCCTGATTGAATCAGTGAATGGAACACAGTTACCT 1942
 1269 rAlaThrGlyThr...IleGlyGlyThrIleSerGlyAsnThrValAsnV 1285
 1943 TCGAA.....CTACTGATAATCGAGCAGGTGCTGAT 1974
 1285 alThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 1301

1975 GCTACTAAG.....GATGACGGTGCTACTCAAG 2003
|||||::: :::::||||
1302 AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 1318
2004 GTATTTCACAACATTATGACACAGCAATCGGTAGATACAGTGTAAAGTCGGG 2053
:ACCTCT:: |||::: :::::||||
1318 rGluAlaSerSerHisIleThrSerAlaLysGlyGlnValAsnLeuSerA 1335
2054 CT.....CTGGGAGGAGTTAAACCGCCACGACGACGAGAGTG 2088
|| |||::: :::::||||
1335 laGlnAspGlySerValAlaGlySerIleAsnAlaAla.....AsnVal 1349
2089 ATACCCACGACAGCTGGAGCACTGTACATACCTGCGCTGGATTGAGAAATGA 2138
::: |||::: |||
1350 ThrLeuAsnThrThrGlyThrLeuThrThr.....ValLysGlySer 1363
2139 TCAAATCAATGGAATCCACCAAGACCTGAAATTAATAAGGATGATGTTTC 2188
:||||::: :::::|||||
1363 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 1380
2189 AA.....CACAGCAAGTGTGTTTCACCAGA 2214
|||||::: |||
1380 luLeuAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn 1396
2215 ACATCTCGGGAGCTCATTTGTGGCTTCGTGATGTCCTCAATGCTCCCAT 2264
:|||||::: |||
1397 AlaAsnGlySerGlySerValIleAlaThrThrSerSerArgValAsnIl 1413
2265 ACCT...GATCTCTTCCCACTGCGCCAAATC.....ACCG 2296
| |||||::: ::::: |||
1413 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleSerLysAsnG 1430
2297 ACCTGAGCGCGAATTCACGGGGGCGAGCTCTCAATTAATCTGATCTGGACA 2346
:||||::: ::::: |||
1430 lyIleAsnThrValLeuLeuLysGlyValLysIleAspValLysTyrIle 1446
2347 GTCCTCTGGG...GATGATTATGACATCGGAACACACTCACAGTATATCAT 2393
|| |||::: ::::: |||
1447 GlnProGlyIleAlaSerValAspGluValIleGluAlaLys..... 1460
2394 TCCAATAAGTACAAGCTATTCTTGATCTCAGACAGCAAGTCTCAATGAATCTC 2443
:||||::: |||||
1461 rGlyIleGluLysValLysAspLeuSerAspGluArgGluAlaL 1477
2444 TT 2445
1477 eu 1477

seq name: /cgg2 6/otodata/2/iaa/5B COMB.pep:US-08-617-697-2

```

seq_documentation_block:
; Sequence 2, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; OF INVENTION: Of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; BUILDING: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697

```

```

FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/302,832
  FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US PCT/US93/02166
  FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
  NAME: Berkstresser, Jerry W
  REGISTRATION NUMBER: 22,651
  REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (703) 415-0810
  TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1536 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
US-08-617-697-2

alignment_scores:
  Quality: 138.50      Length: 936
  Ratio: 0.318        Gaps: 49
  Percent Similarity: 46.581      Percent Identity: 20.406

alignment_block:
US-09-649-696-20 x US-08-617-697-2  ..

Align seg 1/1 to: US-08-617-697-2 from: 1 to: 1536

217 TTTGAAGCTACA.....GGAAAGCGATTTTATTTCAAAATGT 254
631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnII 643
255 TGCATTTTGATTCCTGAACATCGGAAGACAAAGCTGACTATGTGAGAC 304
643 eSerMetValLeuPro.....LysAsnGluSerGlyTyrAspLysP 657
305 CAAACCTTTGAGACCTCAAAAATGCTGATGTTCTG...GTTGCTGAGTCT 351
657 helysGlyArgThrTyrTrpAsnLeuThrSerLeuAsnValSerGluSer 673
352 .....ACTCCTCCAGGTAAATGAT..... 369
674 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspSerAlaGlyTh 690
370 .....GAACCCCTACACT.....GAGCAGA 388
690 rLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThrT 707
389 TGGCAACTGTGGAGAGA.....AGGTTGAAGGATCCACCTCACTCCT 432
707 hrPheAsn.ValGluArgAsnAlaArgValAsnPheAspIleLysAlaPr 723
433 GATTTCATTGCAGGAAAAAAGTTAGCTGAATATTGGACCACCAAGTAGGCG 482
723 oilegIleAsnLysTyrSerSerLeuAsnTyrAlaSerPheAsnGlyA 740
483 ATTTGTCATCAGTGGGCTCATCTACGATGGGGAGTATTTTCACGAGTACA 532
740 snIleSerValSerGly.....GlyGlySerValAspPheThr 752
533 ATAATGATGAGAAATCTACTATTATCCAATGGAAGAATAACA.....AGCA 576
753.LeuLeuAlaSerSerSerAsn.ValGlnThrProGlyValIleAsnS 769
577 GTAAGATGTTTCAGCAGGATATTACTCGGTACAATGTAGTAAAGAGTGTC 626

```

769 erLysTyrPheAsn.valSerThrGlySerSerLeuArgPheLysThrSe 785
627 GGGA.....GGCAGGTGTACACCAAAAGATCCACATCAAT. 563
785 rGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeuAsnA 802
664AAAGTACAGGACTCTATGAAAAGGATGTAGTTGTT 702
802 laThrGlyLysAsnIleThrLeuGlnValGluGlyThrAspGlyMet 818
703 CTCCAAATCCCGCAGCAGGAGAGGCTTCTATAATCTTT..... 741
819 ileGlyLysGlyIleValAlaLysAsnIleThrPheGluGlyGlyAs 835
742GCACAACATGTTGATCTATAGTTG 766
835 nIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsnValT 852
767 AATTCGTACAGAA.....CAAAAC 786
852 hrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPheAspAsn 868
787 CACAACAAAGAGCTCCAAACAGCAAAAT.....CAAAATGCAA 827
869 HisGlnLysProLeuThrIleLysLysAspValIleIleAsnSerGlyAs 885
828 TCTCCGAACCATCGGAAGTG.....ATCC 853
885 nLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuThrValG 902
854 GTGATTCTGAGGACTTTAAGAAACCACTCTATGACA..... 891
902 luSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsnValGly 918
892ACACAGCCAAATCCCACTTCTCATTCCTGCAGATTGG 932
919 GlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLysGlyG 935
933 ACAAGAAATGTTGTGTTAGTCTTCACAAATCTGGAACCATGGCAGCTG 982
935 yAlaArgPheLysAspIleAspAsnSerLysAsnLeuSerIleThrA 952
983 GTAACGCCCTC.....AATCGACTGAATCAA 1008
952 snSerSerThrTyArgThrIleIleSerGlyAsnIleThrAsnLys 968
1009 GCAGGCCAGCTTTCTGCTGCAGCAGATTGAGCTGGGCTCTCGGTTGG 1058
969 AsnGlyAspLeuAsnIleThrAsnGlu.....GlySer. 979
1059 GATGTGCATTTGACAGTGCTGCCATGTACAAAGTGAACATACAG. 1107
980AspThrGluMetGlnIleGlyAspValSerGlnL 992
1108ATAACAGTGCAGTGACAGGACACACTCGCCAAAGA 1146
992 ysGluLysAsnLeuThrIleSerSerAspLysIleAsnIleThrLysGln 1008
1147TTACCTGCAGCAGCTTCAGAGGAGCTCCATCTGCAGCGGCT 1190
1009 IleThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAspAlaTh 1025
1191 TGCATCGCATTTACTGTGATTAGNAGAA..... 1221
1025 rAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnA 1042
1222TATCAACTGATGA 1236
1042 spLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGly 1058
1237 TCTGAATTTGCTGCTGACGAGTGGGAAGACACACTATAAGTGGG... 1284
1059 SerAsp.....LeuThrIleGlyAsnThrAsnSerAlaAspGlyTh 1072

1285TGCTTTAACGAGGTCAAAAGT.....GGTG 1312
1072 rAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089
1313 CCATCATCCACACAGTCGCTTTGGGCCCTCTGCAGCTCAAGAACTAGAG 1362
1089 laAspGlyHisLysValThrLeuHis..... 1097
1263 GAGCTGTCCAAAATG...ACAGAGGTTTACAGACATATGCTTCAGATCA 1409
1098SerLysValGluThrSerGlySerAsnAsnThrGluAspSe 1112
1410 AGTTCAGAAACATGGCTCATTCATGCTTTTGGGCCCTTTTCATCAGGAA 1459
1112 rSerAspAsnAla..... 1117
1460 ATGGAGCTGCTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAACC 1509
1118GlyLeuThrIleAspAlaLysAsnValThr 1127
1510 CTCCAGACAGCAGTGGATGAATGCACAGTATCCTGGACAGCACCGT 1559
1128 ValAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144
1560 GGAAGGACACTTTGTTTCTATCATTACCTGGACAACGACGCTCCCCAAA 1609
1144 rGlyGlu.....IleThrThrLysThr..... 1151
1610 TCCTCTCTGGGATCCAGTGGACAGAAGCTGCTTTGTAGTGGAC 1659
1152GlyThrThrIleAsn 1156
1660 AAAACACCAAAATGGCTTACCTCCAAATCCAGGCACTGCTAAGGTGG 1709
1157 AlaThrThrGly.....AsnValGluIleThrAlaGlnThrG 1169
1710 CACTTGGAAATACAGTCTCAAGCAAGCTCACAACTTGACCTGACTG 1759
1169 ySerIleLeuGlyGlyIleGluSerSerGlySerValThrLeuThrA 1186
1760 TCAGTCCCGTGG.....TCAATGCTACCTGCTCCCAATTACA 1800
1186 laThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
1801 GTGACTTCCAAACGAAGACAGCACACAGC..... 1830
1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLy 1219
1830 1830
1219 sGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyLyt 1236
1831AAATTCGCCAGCCCTCTGGTAGTT 1854
1236 hrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThrThr 1252
1855 TATGCAATATTGCG.....CAAGGAGCTCCCAATTCAG 1892
1253 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 1269
1893 GGCCAGTGTCAAGCCCTGATTGAATCAGTGAATGGAACAGTTACCT 1942
1269 rAlaThrGlyThr...IleGlyGlyThrIleSerGlyAsnThrValAsnV 1285
1943 TGGAA.....CTACTGATAATGAGCAGGTGCTGAT 1974
1285 alThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 1301
1975 GCTACTAAG.....GATGAGGCTGCTACTCAAG 2003
1302 AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 1318

2004 GTATTTTCAACACTATGACACGAATGTTAGATACATAGTGTAAAGTCGGG 2053
1318 rGluAaSerSerHisIleThrSerAlaIysGlyGlnValAsnLeuSerA 1335
2054 CT.....CTGGGAGGAGTTAAACCCAGCAGCAGCAGGAGTG 2088
1335 laGlnAspGlySerValaIagIySerIleAsnAlaA.....AsnVal 1349
2089 ATACCCAGCAGAGTGGAGCAGCTGTACATACCTGGCTGGATTGAGATGA 2138
1350 ThrLeuAsnThrGlyThrLeuThr.....ValIysGlySe 1363
2139 TGAATACAAATGGAATCCACACACCTGAAATTAATAGATGATTC 2188
1363 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaIyAspAlaG 1380
2189 AA.....CACAGCAAGTGTGTTTCAGCAGA 2214
1380 luLeuAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn 1396
2215 ACATCTCGGAGGCTCATTTGTGGCTTCTGATGTCCTCCAAATGCTCCAT 2264
1397 AlaAsnGlySerGlySerValIleAlaThrThrSerArgValAsnII 1413
2265 ACCT...GATCTCTCCACCTGGCCAAATC.....ACCG 2296
1413 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleIleSerIyAsnG 1430
2297 ACCTGAAGCGCGAAATTCACGGGGGCGAGTCTCATTAATCTGACTTGGACA 2346
1430 lyIleAsnThrValLeuLeuIyGlyValIyIleAspValIyTyIle 1446
2347 GCTCTGGG...GATGATTATCACCATGGAACAGCTCACAAATATATCAT 2393
1447 GlnProGlyIleAlaSerValAspGluValIleGluAlaIyS..... 1460
2394 TCGAATAAGTACAGTATTTGATCTCAGACAGCAAGTTCATGAATCTC 2443
1461 ArgIleLeuGluIyValIyAspLeuSerAspGluGluArgGluAlaL 1477
2444 TT 2445
1477 eu 1477

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-719-641-2
seq_documentation_block:
; Sequence 2, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832

FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA: US PCT/US93/02166
APPLICATION NUMBER: 16-MAR-1993
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA: GB 9205704.1
APPLICATION NUMBER: 16-MAR-1992
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
US-08-719-641-2
alignment_scores:
Quality: 138.50 Length: 936
Ratio: 0.318 Gaps: 49
Percent Similarity: 46.581 Percent Identity: 20.406
alignment_block:
US-08-049-696-20 x US-08-719-641-2 ..
Align seg 1/1 to: US-08-719-641-2 from: 1 to: 1536
217 TTTGAAGCTACA.....GGAAGCGATTATTTTCAAAATGT 254
631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnII 643
255 TGCATTTTCATTCCTCAACATGGAAGCAAGGCTGACTATGTGAGAC 304
643 eSerMetValLeuPro.....LysAsnGluSerGlyTyAspLysP 657
305 CAAACCTTGAGACTACAAAATGCTGATGTTCTG...GTTGCTGAGTCT 351
657 helysGlyArgThrTyTipAsnLeuThrSerLeuAsnValSerGluSer 673
352ACTCTCCAGTAATGAT..... 369
674 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspSerAlaGlyTh 690
370GAACCCCTACACT.....GAGCAGA 388
690 rLeuThrGlnProTyAsnLeuAsnGlyIleSerPheAsnLysAspThr 707
389 TGGCAACTGTGGAGAGA.....AGGTGAAAGGATCCACCTCACCTCT 432
707 hrPheAsn.ValGluArgAsnAlaArgValAsnPheAspIleIyAlaPr 723
433 CATTTTCATTCGAGGAAAAAGTTAGCTGAATATGGACCACAAAGTAGGGC 482
723 olleGlyIleAsnLysTySerSerLeuAsnTyAlaSerPheAsnGlyA 740
483 ATTTGCTCATGAGTGGCTCATCTACGATGGGAGATATTTCACGAGTACA 532
740 snIleSerValSerGly.....GlyGlySerValAspPheThr 752
533 ATAATGATGAGAAATTCCTACTTATCCAAATGGAAGAATACA.....AGCA 576
753 leuLeuAlaSerSerSerAsn.ValGlnThrProGlyValValIleAsnS 769
577 GPAAGATGTTACAGAGGATTTACTGGTACAAATGTAGTAAAGAGTGTCA 626
769 erLysTyPheAsn.ValSerThrGlySerLeuArgPheLysThrSe 785

1265TGCTTTAACGAGGTCAACAAGT.....GGTG 1311

1072 rAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089

1313 CCATCATCĈCACACAGTGCCTTTGGGCCCTCTGCAGCTCAAGAAC TAG 1362

1089 laAspGlyHisLysValThrLeuHis..... 1097

1363 GAGCTGTCĈAAATG.....ACAGGAGTTTACAGACATATGCTTCAGATCA 1409

1098SerLysValGluThrSerGlySerAsnAsnThrGluAspSe 1112

1410 AGTTCAGAACAAATGGCCTCAITGATGCTTTGGGCCCTTTCATCAGGAA 1459

1112 rSerAspAsnAla..... 1117

1460 ATGAGAGTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAAC 1509

1118GlyLeuThrIleAspAlaLysAsnValThr 1127

1510 CTCAGAACACGCCAGTGAATGGCACAGTATGCTGGAGCAGCACCTG 1559

1126 ValAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144

1560 GGGAAAGGACACTTGTTCCTATCACCTGGACACGCGACCTCCCCAAA 1609

1144 rGlyGlu.....IleThrLysThr..... 1151

1610 TCCTTCTCTGGGATCCAGTGGACAGAAAGGTGGCTTCTAGTGGAC 1659

1152GlyThrThrIleAsn 1156

1660 AAAAACACĈAAATGGCTACTCTCAAAATCCAGCATGCTGAAGTTGG 1709

1157 AlaThrThrGly.....AsnValGluIleThrAlaGlnThrG 1169

1710 CACTTGGAAATACAGTCTGCAAGCAAGCTCAACAACCTTGACCTGCAGTG 1759

1169 ySerIleLeuGlyIleGluSerSerGlySerValThrLeuThrA 1186

1760 TCAGTCCCGTGG.....TCCAATGCTACCTGCCTCCAATTACA 1800

1186 laThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202

1801 GTGACTTCCĈAAACGAACAGGACACCAGC..... 1830

1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLy 1219

1830 1830

1219 sGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyGlyT 1236

1831AAATCCCCAGCCCTCTGGTAGTT 1854

1236 hrIleSerGlyClyThrValGluValLysAlaThrGluSerLeuThrThr 1252

1855 TATGCAAAATATCGC.....CAAGGAGCCTCCCCAATTCTCAG 1892

1253 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 1269

1893 GGCCAGTGTCAACGCCCTGATTGAATCAGTGAATGGAAACACAGTTACCT 1942

1269 rAlaThrGlyThr...IleGlyGlyThrIleSerGlyAsnThrValAsn 1285

1943 TGGAA.....CTACTGGATAATGGACGAGTCTGAT 1974

1285 alThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 1301

1375 GCTACTAAG.....GATGACGGTGTCTACTCAAG 2003

1302 AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 1318

2004 GTATTTCCAACTTATGACACGAATGGTAGATACAGTGTAAAGTGC GCG 2053

us-09-049-696-20.ra1

Tue Apr 2 09:39:58 2002

```

1318 rGluAlaSerHisIleThrSerAlaLysGlyGlnValAsnLeuSerA 1335
2054 CT.....CTGGGAGGAGTTAAACGACGACGACGAGAGTG 2088
1335 laGlnAspGlySerValAlaGlySerIleAsnAlaAla.....AsnVal 1349
2089 ATACCCAGAGAGTGGAGCAGCTGTACATACCTGGCTGGATTGAGATGA 2138
1350 ThrLeuAsnThrThrGlyThrLeuThr.....ValLysGlySe 1363
2139 TGAATACAAATGAATCCACCAAGACCTGAATTAATAAGATGATGTC 2188
1363 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 1380
2189 AA.....CACAAAGCAAGTGTTCACGAGA 2214
1380 LeuAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn 1396
2215 ACATCTCGGAGGCTATTGTGCTTCGTATGTCCTCCAAATGCTCCAT 2264
1397 AlaAsnGlySerGlySerValIleAlaThrThrSerSerArgValAsnIl 1413
2265 ACCT....GATCTCTCCACCTGCGCCAAATC.....ACCG 2296
1413 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleIleSerLysAsnG 1430
2297 ACCTGAAGCGGAATTCACGGGCGAGCTCTATTATCTGACCTGGACA 2346
1430 lYleAsnThrValLeuLeuLysGlyValLysIleAspValLysTyrIle 1446
2347 GCTCTGGG...GATGATTATGACCATGGAACAGCTCACAAGTATATCAT 2393
1447 GlnProGlyIleAlaSerValAspGluValIleGluAlaLys..... 1460
2394 TCGAATAGTACAGATATCTTGATCTCAGACAGCAAGTTCAATGAATCTC 2443
1461 ArgIleLeuGluLysValLysAspLeuSerAspGluArgGluAlaL 1477
2444 TT 2445
1477 eu 1477

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.US-08-617-697-10

seq_documentation_block:
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1600 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-10

alignment_scores:
Quality: 138.50 Length: 680
Ratio: 0.425 Caps: 28
Percent Similarity: 47.941 Percent Identity: 20.882

alignment_block:
US-09-049-696-20 x US-08-617-697-10 ..
Align seg 1/1 to: US-08-617-697-10 from: 1 to: 1600

865 GACITTTAAGAAAACCACTCTCTATGACACACAGCCACCAATCCCACTT 914
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 AspGlyAsnLysThrThrIleArgAsnSerValAsnAlaIleAsnTr 102
913 CTCATTGCTGCAGATTGGACAAAGAAATTCGTGTTTAGTCTTGCACAAAT 964
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
102 pLysGlnPheAsnIleAspGlnAsnGluMetGluGlnPheLeuGlnGluS 119
963 CTGAAGACATGGCGACTGTTAACCCTCAATCGA..... 999
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
113 erSerAsnSerAlaValPheAsnArgValThrSerAspGlnIleSerGln 135
1000 .....CTGAATCAAGCAGCGGCTTTCTCTGTCGAG..... 1032
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
135 LeuLysGlyIleLeuAspSerAsnGlyGlnValPheLeuIleAsnProAs 152
1033 .....ACAGTTGAGTGGGTCCTGGTGGGATGGTGACATTGAC.. 1074
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
152 nGlyIleThrIleGlyLysAspAlaIleIleAsnThrAsnGlyPheThrA 169
1075 ..AGTGTGCCCATGTACAAAGTGAACACTCATACAGATAAACAGTGGCAGT 1122
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
169 laSerThrLeuAspIleSerAsnGluAsnIleLysAlaArgAsnPheThr 185
1123 GACAGGACACACTCGCCAAAGAAATTCCTGCAGCAGCTTCAGGAGGAC 1172
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
185 LeuGluGlnThrLysAspLysAlaLeuAlaGluIleValAsnHisGlyLe 202
1173 CTCATCTGCAGCGGCTTCGATCGGCAATTTACTGTGATTAGGAGAAAT 1222
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
202 u...IleThrValGlyLysAspGlySerValAsnLeuIleGlyLysV 218
1223 ATCCAACCTGAT.....GGATCTGAAATTTGCTGCTGTG 1254
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
218 alLysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeuLeu 234
1253 ACGGATGGGGAAGACACACATATAGTGGTGTCTTAAACAGGTCAAACA 1304
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
235 Ala...GlyGlnLysIleThrIleSerAspIleIleAsnPro..... 247
1305 AAGTGTGCCATTCACACAGCTCGCTTTGGGGCCCTCTCCAGCTCAAG 1354
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
248 .....ThrIleThrTyrSerIleAlaAlaProGluAsnGluAlaIleA 262

```

1355 AACTAGAGGAGCTGCCAAATGACAGGAGGTTTACAGACATATGCTTCA 1404
262 snLeuGlyAsnIlePheAlaLysGlyGlyAsnIleAsnValArgAlaAla 278
1405 GATCAAGTTCAGAACAAATGGCTCAATT.....GATGCTTTTGGGGCCCT 1448
279 ...ThrIleArgAsnLysGlyLysLeuSerAlaAspSerValSerLysAs 294
1449 TTCATCAGAAATGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTA 1498
294 pLysSerGlyAsnIleValLeuSerAlaLysGlyGlyAlaGluIleG 311
1499 AGGGA...TTAACCTCCAGAACCGCATGGATGAATGGCAGATGATC 1545
311 lyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGlyLysLeuMet 327
1546 GTG..... 1548
328 IleThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLeuSe 344
1549GACAGCAGGTGG 1561
344 rGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGluArgGlyGluG 361
1562 GAAAGGACACTTGTCTTATCCTGACGACGACGAGCTCCCAATC 1611
361 lyLysAsnGlyIleGlnLeuAlaLysLysThrLeuGluLysGlySer 377
1612 CTTCTCTGGATCCGAGTCACGACAGCAAGGTGGC..... 1647
378 ThrIle...AsnValSerGlyLysGluLysGlyGlyArgAlaIleValTr 393
1648TTGTAGTGGACAAAACACCAAAATGGCTACCTCCAAA 1687
393 pGlyAspIleAlaLeuIleAspGlyAsnIleAsn.....AlaGlnG 407
1688 TCCAGGAGTCTTAAGGTGCTGCTGCAATACAGCTGCAAGCAAGC 1737
407 lySerAspIleAlaLysThrGlyGlyPhe.....ValGluThrSer 420
1738 TCACAAACCTTGACCCCTG..... 1755
421 GlyHisAspLeuSerIleGlyAspValIleValAspAlaLysGluTr 437
1756ACTGTCAGTCCGCTGCT 1774
437 pLeuLeuAspProAspAspValSerIleGluThrLeuThrSerGlyArgA 454
1775 CCAATGCTACCTGCTCCCAATTACAGTGACTTCCAAAACGACAGGAC 1824
454 snAsnThrGlyGluAsnGlnGlyTyrThrGlyAspGlyThrLysGlu 470
1825 ACCAGCAAA.....TTCCCGACCCCTCGGTAGTTTATGCAATAT 1865
471 SerProLysGlyAsnSerIleSerLysProThrLeuThrAsnSerThrLe 487
1866 TCGCAAGAGGAGCTCCCAATTCCTCAGGCGCAGTGTACAGCCCGATG 1915
487 uGluGln.....IleLeuArgArgGlySerTyrValAsnIleT 500
1916 ATCAGTGAATGGAACAGCTTACCTTGAACATA...CTGATAATGA 1962
500 hrAlaAsnAsnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGly 516
1963 GCAGGTGCTGCTACTACTAAGGATGACGGTGTCTACTCAAGGTATTTCAC 2012
517 SerLeuThrLeuHisThrLysArgAspGlyValLysIleAsnGlyAspI 533
2013 AACTTATGACCAATGTTAGATACAGTGTAAAGTCCGGGCTCGGAG 2062
533 eThrSerAsnGluAsnGlyAsnLeuThrIleLysAlaGlySerTrpValA 550
2063 GAGTTAACGCCACGAGAGGAGTGTATACCCAGCAGAGTGAGGACTG 2112

550 spValHis.....LysAsnIleThrLeuGlyThrGlyPheLeu 562
2113 TACATACCTCGCTGGATGAGATGAAATACAATGGAATCCACCAAG 2162
563 AsnIle.....ValAlaGlyAspSerValAlaPheGluArgGluG 576
2163 ACCTGAA..... 2169
576 yAspLysAlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrI 593
2170ATTAATAAGGATGATGTTCAACAAG.....CAAGGTGTTTC 2208
593 leThrValAsnLysAspLysGlnPheArgPheAsnAsnValSerLeu 609
2209 AGCAGAACATCCTCGGAGGCTCATTTGTGGCTTCTGCTGCCAAATGC 2258
610 AsnGlyThrGlyLysGlyLeuLysPheIleAlaAsnGln..... 622
2259 TCCATACCTGATCTCTCCACCTGGCCAAATCACC.....GACC 2299
623AsnAsnPheThrHisLysPheAspG 631
2300 TGAAGCGGAAATTCACGGGCGAGTCTCATTAATCTGACT..... 2340
631 lyGluIleAsnIleSerGlyIleValThrIleAsnGlnThrThrLysLys 647
2341TGGACAGCTCTCGGGGATGATATGACCAT..... 2370
648 AspValLysTyrTrpAsnAlaSerLysAspSerTyrTrpAsnValSerSe 664
2371GGACAGCTCACAGTAT.....ATCATTCGAA 2398
664 rLeuThrLeuAsnThrValGlnLysPheThrPheIleLysPheValAsp 681
2399 TAAGTACAAGTATCTTGATCTCAGAGCAAGTTCAATGAATCTCTCAA 2448
681 eGlySerAsnGlyGlnAspLeuArgSerArgArgSerPheAlaGly 697
2449 GTCAATFACTGCTCTCATCCCAAGAA.....GCCAA 2483
698 ValHisPheAsnGlyIleGlyGlyLysThrAsnPheAsnIleGlyAlaAs 714
2484 CTCTGAGGAGTCTTTTGTTTAAACCAGAAACATTACT 2523
714 nAlaLysAlaLeuPheLysLeuLysProAsnAlaAlaThr 727
seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-296-791-6
seq_documentation_block:
Sequence 6, Application US/08296791
Patent No. 6245337
GENERAL INFORMATION:
APPLICANT: St. Gene III, Joseph W.
APPLICANT: Falkow, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELEPHONE: (415) 398-3249
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; US-08-296-791-6

alignment_scores:
  Quality: 130.00      Length: 956
  Ratio: 0.332        Gaps: 53
  Percent Similarity: 41.004      Percent Identity: 21.234

alignment_block:
US-09-049-696-20 x US-08-296-791-6 ..
Align seg 1/1 to: US-08-296-791-6 from: 1 to: 1848

349 TCTACTCTCCAGGT.....AATGATGAACCTACACTGAGCAG..... 387
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
278 SerAspProLysGlyLeuSerGlnAspProLeuThrAsnTyrAlaVala 294
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
388 .ATGGGCAACTGGA.....GAGAAAGGTG 412
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
294 lLeuGlyAspSerGlySerProLeuPheValTyrAspArgGluLysGlyL 311
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
413 AAAGATCCACCTC...ACTCCTGATTTTCATTCGACGAAAGAGTAGCT 459
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
311 ysTrpLeuPheLeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLys 327
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
460 GAATATGGACCAAGGATGAGTCATTGTCATGAGTGGGCTCATCTACG 509
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 SerTrp...GlnGluTrpAsnIleTyrLysHisGluPheAla..... 340
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
510 ATGGGGAGTATTGACGAGTACAAATGATGAGAAATCTACTTATCCA 559
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
341 .....GluLysIleTyr..... 344
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
560 ATGGAAGAATACAAGCAGTAAGATGTTACGAGGT...ATTACTGGTACA 606
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345 .....GlnGlnTyrSerAlaGlySerLeuIleGlySer 355
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
607 AATGTAGTAAGAAGTGTGAGGAGGCGAGCTGTTACACCAAAAGATGCAC 656
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
356 Asn.....ThrGlnTyrTrpGlnAlaTh 364
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
657 A.....TTCAATAAAGTACAGGAGCTCTATGAAAAAGGATGTGAGTTTG 700
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 rGlySerThrSerThrIleThrGly..... 372
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
701 TTCTCCAATCCCGCAGCAGGAGAGGCTTCTATATGTTTGCACAAAT 750
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
373 .....GlyGlyGluProLeuSerValAspLeuThrAspGly 384
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
751 GTTGATTTCTATAGTTGAATCTGTACAGAACAAACACCAACAAAGAACG 800
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
385 LysAsp.....LysProAsnHisGlyLysSerIl 394
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
801 TCCAAACAAAGCAAAATCAAAATGCAATCTCCGA..... 834
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
394 eThrLeuLysGlySerGlyThrLeuThrLeuAsnAsnHisIleAspGlnG 411
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
834 ..... 834

```


1546 GTG.....GACAGCAGCGTGGGAAGAGACACTTGTCTTCTTATCACCTG 1589
::: ||| ::||| ||| ||||| |||||
700 LeuAsnGlyAspLeuAsnValGluLysGlyThrLeuPheLeuSerGlyAr 716
1590 GACAAACGAGCGTCCCAATCTCTCTGGGATCCAGTGGACACAGAGC 1639
||| |||::: ||| ||| ::|||::: |||
716 gProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysA 733
1640 AAGTGGCTTTGTA.....GTG 1656
:: |||::: |||
733 spProHisPheThrGluAsnAsnGluValValGluAspAspTrpIle 749
1657 GACAAAACACCAAAATGGCTACCTCCAAATCCAGGCATTGCTAAGGT 1706
::: |||::: ||| ::: |||::: ||| |||
750 AsnArgAsnPheLysAlaThrThrMetAsnValThrGlyAsnAla..... 764
1707 TGGCCTTGGAAATACAGTCTCAAGCAGAGCTCACAACCTTGACCTGA 1756
||| |||::: |||::: |||::: |||::: |||
765SerLeuTyrSerGlyArgAsnValAlaAsnIleT 776
1757 CTGTCAGCTCCGTCGGTCCCAAT.....GCT 1782
|| ::: ||||| |||
776 hrSerAsnIleThrAlaSerAsnAlaGlnValHisIleGlyTyrLys 792
1783 ACCCTGCCTCCAAATACAGTGACTTCCAAA..... 1812
||| ::: ||| |||
793 ThrGlyAspThrValCysValArgSerAspTyrThrGlyTyrValThrCy 809
1813ACGAACAAGCAGACACCAAAATTC...CCAGCC 1843
809 shisAsnSerAsnLeuSerGluLysAlaLeuAsnSerPheAsnProThra 826
1844 CTCTGGTAGTTATGCAATATCCGCCAGGAGCGCTCCCAATTC... 1890
||| ::|||::: |||::: |||::: |||
826 snLeuArgGlyAsnValAsnLeuThrGluAsnAlaSerPheThrLeuGly 842
1891 AGGGCCAGTGTCCACAGCCCTGATGAATCAGTGAATGGAAAAACAGTTAC 1940
::: |||::: |||::: |||::: |||::: |||::: |||::: |||
843 LysAlaAsnLeuPheGlyThrIleGlnSerIleGlyThrSerGlnValAs 859
1941 CTTGGAA..... 1947
::|||:::
859 nLeuLysGluAsnSerHisTrpHisLeuThrGlyAsnSerAsnValAsnG 876
1948CTACTGGATATGGA.....GCAGGTGCTGATGCT 1977
||| ||||| ||| ||||| |||||
876 InLeuAsnLeuThrAsnGlyHisIleHisLeuAsnAlaGlnAsnAspAla 892
1978 ACTAAGGATGACGGTCTCTACTCAAGGTATTTCACAACCTTATGACACGAA 2027
::: ||| |||||::: |||
893 AsnLys.....ValThrThrTyrAsnThr... 900
2028 TGGTAGATACAGTGAAGTGGGCTCTGGGAGGAGTTAACCGAGCCA 2077
::: |||::: |||::: |||::: |||
901LeuThrValAsnSerLeuSerGly..... 908
2078 GACGGAGAGTGATACCCAGCAGAGTGGACACTGTATACATACCTGGCTGG 2127
::: |||::: |||::: |||
909AsnGlySerPheTyrTyr.....Trp 915
2128 ATTGAGATGATGAATACAATGGAATCCACCAAGACCTGAAATTAATAA 2177
::: |||::: |||::: |||
916 ValAspPheThrAsnAsnLysSerAsn.....LysValValValAsnLys 930
2178 GGATGATGTTCAACACACAAACAGTGTGTTTCACAGACATCTCGGGAG 2227
|||
930 s.....SerAlaThrG 934
2228 GCTCATTTGTGCTTCTGTATGCCCAAAATGCTCCCATACCTGATCTCTTC 2277
|||::|||:::
934 LysAsnPheThrLeu..... 938

2278 CCACCTGGCCAAATCACCGACCTGAAGCGGGAATTCACGGGGCGAGCTCT 2327
|||::: |||::: |||::: |||::: |||
939GlnValAlaAspLysThrGlyGluProAsnHisAsnGluLe 952
2328 CATT.....AATCTGACTTGGACAG 2347
||| ||||| |||
952 uThrLeuPheAspAlaSerAsnAlaThrArgAsnAsnLeuGluValThrL 969
2349 CTCCTGGGGATGATTATGACCATGGAAACAGCTCACAAAGTATATCATCGA 2397
::: |||::: |||::: ||| ||||| |||||
969 euAlaAsnGlySerValAspArgGly...AlaTrpLysTyrLysLeuArg 984
2398 ATAAGTACAAGTATTCTTGTATCTCAGA.....GACAAGTTCAA 2435
::: |||::: ||||| |||
985 AsnValAsnGlyArgTyrAspLeuTyrAsnProGluValGluLysArgAs 1001
2435 TGAATCTCTTCAAGTGAATACTACTGCTCTCATCCCCAAAGGAAGCCAAC 2485
|||::: |||::: ||||| |||
1001 nGlnThr.....ValAspThrThr..... 1007
2485 CTGAGGAAGCTCTTTTGTGTTTAAACCAAGAAACATTAATCTTTGAAATGGC 2535
||| ||||| |||
1009AsnIleThrThrProAsnAsp 1014
2535 ACAGATCTTTTCATTCTTATTCAGGCTGTGATAAGGTGATCTGAAATC 2585
||| ||||| ||| |||::: |||::: |||
1015IleGlnAla...AspAlaProSerAlaGlnSe 1024
2585 AGAAATATCCAAATTCAGAGTATCTTTGTTTATTTCTCTCCACAGACTC 2635
|||::: |||::: |||::: |||::: |||::: |||
1024 rAsnAsnGluGluIleAlaArgValGluThrProValProProAlap 1041
2635 CG.....CCAGAGACACCTAGTCCT 2655
|| ||| ||||| |||
1041 roAlaThrGluSerAlaIleAlaSerGluGlnProGluThr...ArgPro 1056
2655 GATGAAACGCTCTGCTCCT 2673
||| |||::: |||
1057 AlaGluThrAlaGlnPro 1062

us-09-049-696-20.ra

Tue Apr 2 09:39:58 2002

This Page Blank (uspto)

OM of: US-09-049-696-20 to: A_Geneseq_1101.* out_format : pfs

Date: Mar 30, 2002 2:28 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL-frame2_n2p.model -DEV=rlp
-Q/cgn2_1/USPTO.spool/US09049696/runat_28032002_145237_2001/app_query.fasta_1.12579
-DB=A_Geneseq_1101 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GCAPOP=4.000 -CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOCLALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1_293 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-049-696-20
Query length: 2983
Database: A_Geneseq_1101.*
Database sequences: 522463
Database length: 74073290
Search time (sec): 819.910000
```

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AA11789 +		199.50	322.20	6.1e-10	92
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA03643 +		193.50	312.34	2.2e-09	91
/SID22/gcgdata/geneseq/geneseq/AA1997.DAT:AAW30292 +		143.50	193.56	0.0005	1601
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA03643 +		143.00	187.66	0.0007	2411
/SID22/gcgdata/geneseq/geneseq/AA1996.DAT:AA03643 +		141.50	185.47	0.0010	2353
/SID22/gcgdata/geneseq/geneseq/AA1993.DAT:AA03643 +		140.50	189.13	0.0009	1529
/SID22/gcgdata/geneseq/geneseq/AA1998.DAT:AAW30292 +		139.50	187.41	0.0012	1536
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA03643 +		139.50	183.91	0.0014	2039
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA03643 +		138.50	189.93	0.0012	1095
/SID22/gcgdata/geneseq/geneseq/AA1993.DAT:AA03643 +		138.50	185.75	0.0015	1536
/SID22/gcgdata/geneseq/geneseq/AA1994.DAT:AA03643 +		138.50	185.75	0.0015	1536
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA03643 +		138.50	185.75	0.0015	1536
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AA03643 +		135.50	184.31	0.0023	1152
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AAU05396 +		135.50	184.31	0.0023	1152
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AAU05396 +		134.00	142.81	0.0205	26926
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AAU05396 +		133.50	177.15	0.0043	1569
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA01834 +		133.00	180.70	0.0039	1101
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AAE01422 +		132.00	179.98	0.0046	1020
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AA03643 +		132.00	161.92	0.0108	4393
/SID22/gcgdata/geneseq/geneseq/AA1997.DAT:AAW30293 +		131.50	174.08	0.0065	1536

seq_name: /SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AA11789

seq_documentation_block:

ID: AA11789 standard; Protein; 92 AA.

XX AC AA11789;

XX DT 18-JUN-1999 (first entry)

XX XX Human 5' EST secreted protein SEQ ID No: 389.

Human; secreted protein; EST: expressed sequence tag; diagnosis:
forensic; gene therapy; chromosome mapping; signal peptide; prostate;
upstream regulatory sequence; cytokine activity; cell proliferation;
differentiation; haematopoiesis regulation; tissue growth regulation;
reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
thrombolytic; anti-inflammatory; tumour inhibition.

OS Homo sapiens.

```
XX WC3906550-A2.
XX PD 11-FEB-1999.
XX PF 31-JUL-1998; 98WO-IB01232.
XX PR 01-AUG-1997; 97US-0905144.
XX PA (GSET ) GENSET.
XX PI Ductert A, Dumas Milne Edwards J, Lacroix B;
XX DR WPI: 1999-153780/13.
XX DR N-PSDB; AAX40511.
XX FT New isolated prostate-derived nucleic acids - used to develop
XX PT products which may have cytokine, immune regulatory, haematopoiesis
XX PT regulating, anti-inflammatory or tumour inhibition activity
XX PS Claim 34; Page 546; 675pp; English.
XX CC AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for
XX CC human secreted proteins expressed in prostate, and encode the proteins
XX CC given in AA11716 to AA11993 respectively. The proteins given represent
XX CC the signal peptide and an N-terminal fragment of a secreted protein. The
XX CC nucleic acid sequences can be used for producing secreted human gene
XX CC products. The proteins obtained may have cytokine activity, cell
XX CC proliferation and differentiation activity, haematopoiesis regulating
XX CC activity, tissue growth regulating activity, reproductive hormone
XX CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
XX CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX CC activity, tumour inhibition activity or other activities. The products
XX CC can be used in forensic, gene therapy and chromosome mapping procedures.
XX CC The sequences can also be used for obtaining corresponding promoter
XX CC sequences. The nucleic acids encoding the signal peptides can be used for
XX CC directing extracellular secretion of a polypeptide or the insertion of a
XX CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX SQ Sequence 92 AA;
alignment_scores:
Quality: 199.50 Length: 95
Ratio: 2.733 Gaps: 3
Percent Similarity: 76.842 Percent Identity: 45.263
alignment_block:
US-09:049-696-20 x AA11789 ..
Align seg 1/1 to: AA11789 from: 1 to: 92
4 ATCACAGGAGATGTACAGCAATGGGGCCATTAAAGAGTTCTGTGTCAT 53
1 MetThrGlnArgSerIleAla...GlyProIleCysAsnLeuLysPheVa 16
54 CTTGATTCTTCCCTTCTAGAGGGGCCCTGAGTATTCACCTC..... 96
16 lThrLeuLeu.....ValAlaLeuSerSerGluLeuProPheL 29
97 .....ATTAGCTGAACACATGGCTGTATGAAGGCATTGTCGT 135
29 euGlyAlaGlyValGlnLeuGlnAspAsnGlyTyrAsnGlyLeuLeu 45
136 GCAATCGACCCCAATGTCGACAGATGAACACATTCATCAACATAAA 185
46 AlaIleAsnProGlnValProGluAsnGlnAsnLeuLeuSerAsnIle 62
186 GGACATGTGTGACCCAGGACATCTCTGATCTGTTTGAAGCTACAGGAAAG 235
62 sGluMetIleThrGluAlaSerPheTyrLeuPheAsnAlaThrLysArg 79
```

```
236 GATTTATTTCAAAATGTTGCCATTTGATTCCT 270
|| :||||:||||: ||| ||| |||
79 rgValPhePheArgAsnIleLeuIlePro 90

seq_name: /SIDS2/gcgdata/geneseq/AA2000.DAT:AA03643

seq_documentation_block:
ID AAG03643 standard; Protein; 91 AA.
XX
AC AAG03643;
XX
DT 06-OCT-2000 (first entry)
DE Human secreted protein, SEQ ID NO: 7724.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR N-PSDB; AAC03649.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 13; SEQ ID 7724; 71pp + CD-ROM; English.
XX
XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
SQ Sequence 91 AA;

alignment_scores:
Quality: 193.50 Length: 94
Ratio: 2.688 Gaps: 3
Percent Similarity: 76.596 Percent Identity: 44.681

alignment_block:
US-09-049-696-20 x AAG03643 ..
Align seg 1/1 to: AAG03643 from: 1 to: 91
4 ATCACGGGAGATCTACGCAATGGGCGCATTTAAGAGTTCTGTTGTCAT 53
:||||| ||| ||| ||| ||| ||| ||| ||| |||
1 MetThrGlnArgSerIleAla...GlyProIleCysAsnLeuLysPheVa 16
54 CTTGATTTCTTACCCTTCTAGAGGGGCCCTGAGT.....A 88
```

```
16 jThr.....LeuLeu***Gly*****Ser***LeuPro***LeuG 30
89 ATTCACTCATTCAGCTGAACAACAAATGGCTATGAAGCGCATTTGCTTGCA 138
30 lyAlaGlyValGlnLeuGlnAsnGlyTyrAsnGlyLeuLeuIleAla 46
139 ATCGACCCCAATGTGCCAAGATGAACACTCATTCACAAATAAAGGA 188
47 lIeAsnProGlnValProGluAsnGlnAsnLeuIleSerAsnIleLysG 63
189 CATGTCACCCAGGCATCTCTATCTCTTTTGAAGCTACAGCAAGCGAT 238
63 uMetIleThrGluAlaSerPheTyrLeuPheAsnAlaThrLysArgV 80
239 TTTATTCAAAATGTTGCCATTTTGAATTCCT 270
80 alPhePheArgAsnIleLysIleLeuIlePro 90

seq_name: /SIDS2/gcgdata/geneseq/AA1997.DAT:AAW30292

seq_documentation_block:
ID AAW30292 standard; Protein; 1601 AA.
XX
AC AAW30292;
XX
DT 14-APR-1998 (first entry)
DE Non-typeable Haemophilus high mol.wt. surface protein HMW4.
XX
KW Non-typeable Haemophilus; high molecular weight surface protein;
KW HMW4; immunogen; vaccine; otitis media.
XX
OS Haemophilus influenzae strain 5.
XX
FH Key Location/Qualifiers
FT Misc-difference 372 /note= "encoded by TCT"
FT Misc-difference 400 /note= "encoded by AAT"
FT
XX WO9736914-A1.
XX 09-OCT-1997.
XX 01-APR-1997; 97WO-US04707.
XX 01-APR-1996; 96US-0617697.
XX (BARE/) BARENKAMP S J.
XX Barenkamp SJ;
XX WPI; 1997-503038/46.
XX N-PSDB; AAT90993.
XX
XX High molecular weight proteins of non-typeable Haemophilus
XX influenzae - useful for vaccine production
XX
XX Claim 1; Page 97-102; 183pp; English.
XX
XX This protein comprises the high molecular weight surface protein
CC HMW4 (123 kDa) of non-typeable Haemophilus influenzae strain 5 that
CC has the immunological ability to protect against disease caused by
CC a non-typeable Haemophilus strain and is characterised by at least
CC one surface-exposed B-cell epitope that is recognised by monoclonal
CC antibody AD6. The HMW4 amino acid sequence was deduced from an
CC isolated HMW4 gene (see AAT90993). HMW1 (see AAW30293), HMW2 (see
CC AAW30294) and HMW3 (see AAW30291) have also been identified. A
CC conjugate comprising HMW4 linked to an antigen, hapten or
CC polysaccharide, and a synthetic peptide of 6-150 amino acids
CC corresponding to at least protective epitope of HMW4 are also
CC claimed. HMW proteins, conjugates and peptides can be used in
```

CC vaccines, as immunogens for preparation of antibodies and as
 CC antigens for detection of these antibodies.

XX
 SQ Sequence 1601 AA;

alignment_scores:

Quality: 143.50 Length: 680
 Ratio: 0.440 Gaps: 29
 Percent Similarity: 47.941 Percent Identity: 20.882

alignment_block:

US-09-049-696-20 x AA030292

Align seg 1/1 to: AA030292 from: 1 to: 1601

```

865 GACTTTAAGAAACCACTCTCTATGACACACAGCCACCAATCCACCTT 914
    ||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
86 AspGlyAsnLysThrThrIleArgAsnSerValAsnAlaIleAsnTr 102
915 CTCATTGCTGCAGATTGGCAAGAAATTGCTGTTTAGTCTTGAACAAT 964
    :|||:||||| :|||:||||| :|||:||||| :|||:|||||
102 pLysGlnPheAsnIleAspGlnAsnGluMetValGlnPheLeuGlnGlu 119
965 CTGGAAGCATGGCGACTGTTAACCCTCAATCGA..... 999
|||:||||| :|||:||||| :|||:||||| :|||:|||||
119 erSerAsnSerAlaValPheAsnArgValThrSerAspGlnIleSerGln 135
1000 .....CTGAATCAAGCAGCCAGCTTTCTCTGTCGAG..... 1032
|||:||||| :|||:||||| :|||:||||| :|||:|||||
136 LeuLysGlyIleLeuAspSerAsnGlyGlnValPheLeuIleAsnProAs 152
1033 .....ACAGTTGAGCTGGGTCCTGGGTGGGATGGTGACATTGAC. 1074
|||:||||| :|||:||||| :|||:||||| :|||:|||||
152 nGlyIleThrIleGlyLysAspAlaIleAsnThrAsnGlyPheThra 169
1075 ..AGTCTGCCCATGTACAAGTGAACATCATACAGATAACAGTCGCAGT 1122
|||:||||| :|||:||||| :|||:||||| :|||:|||||
169 laSerThrLeuAspIleSerAsnGluAsnIleLysAlaArgAsnPheThr 185
1123 GACAGGACACACTCGCCAAAGATTACCTGCAGCAGCTTCAGAGGGGAC 1172
:||||| :|||:||||| :|||:||||| :|||:|||||
186 LeuGlnGlnThrLysAspLysAlaLeuAlaGluIleValAsnHisGlyLe 202
1173 GTCCATCTGCAGCGGCTCGATGCGCATTTACTGTGATTAGGAAGAAAT 1222
|||:||||| :|||:||||| :|||:||||| :|||:|||||
202 u...IleThrValGlyLysAspGlySerValAsnLeuIleGlyLysV 218
1223 ATCCAACATGAT.....GGATCTGAATTTGCTGCTG 1254
:||||| :|||:||||| :|||:||||| :|||:|||||
218 alLysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeuLeu 234
1255 ACGGATGGGGAAGACACACTATAAGTGGGTGCTTTAAGCAGGCTAAACA 1304
:||||| :|||:||||| :|||:||||| :|||:|||||
235 Ala...GlyGlnLysIleThrIleSerAspIleIleAsnPro..... 247
1305 AAGTGTGCTCATCATCCACAGCTCGCTTTGGGCGCTCTGCGACTCAAG 1354
:||||| :|||:||||| :|||:||||| :|||:|||||
248 .....ThrIleThrTy+SerIleAlaAlaProGluAsnGluAlaIleA 262
1355 AACTAGAGGAGCTGCATAAATCACAGAGGTTTACAGACATATGCTTCA 1404
:||||| :|||:||||| :|||:||||| :|||:|||||
262 snLeuGlyAspIlePheAlaLysGlyIleAsnIleAsnValArgAlaAla 278
1405 GATCAAGTTTCAGAACCAATGCCTCATT.....GATGCTTTTGGGCGCCT 1448
:||||| :|||:||||| :|||:||||| :|||:|||||
279 ...ThrIleArgAsnLysGlyLysLeuSerAlaAspSerValSerLysAs 294
1449 TTCATCAGGAATGGAGTCTCTCTCAGCGCTCCATCCAGCTTGCAGAGTA 1498
:||||| :|||:||||| :|||:||||| :|||:|||||
294 pLysSerGlyAsnIleValLeuSerAlaLysGluGlyGluAlaGluIleG 311
1499 AGGGA...TTAACCTCCAGACAGCAGTGGTGAATGATGCCAGCTGATC 1545
```

```

||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
311 lyGlyValIleSerAlaGlnAsnGlnAlaLysGlyGlyLysLeuMet 327
1546 GTG..... 1548
328 IleThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLeuSe 344
1549 .....GACAGCACCCTGG 1561
344 rGlyLysGluGlyGlyGluThrTrpLeuGlyGlyAspGluArgGlyGluG 361
1562 GAAAGCAGCACTTTGTTTCTATCACCTGGACACAGCAGCCTCCCCAATC 1611
|||:||||| :|||:||||| :|||:||||| :|||:|||||
361 lyLysAsnGlyIleGlnLeuAlaLysLysThrThrLeuGluLysGlySer 377
1612 CTTCTCTGGGATCCCATGGTGACAGCAAGAGTGGC..... 1647
:||||| :|||:||||| :|||:||||| :|||:|||||
378 ThrIle...AsnValSerGlyLysGluLysGlyGlyArgAlaIleValTr 393
1648 .....TTGTAGTGGCAAAACACCAAAATGGCCTACCTCCAAA 1687
:||||| :|||:||||| :|||:||||| :|||:|||||
393 pGlyAspIleAlaLeuIleAspGlyAsnIleAsn.....AlaGlnG 407
1683 TCCAGGCAATTGCTAAGTTGGCAGCTTGGAAATACACTCTGCAAGCAAGC 1737
|||:||||| :|||:||||| :|||:||||| :|||:|||||
407 lySerAspIleAlaLysThrGlyGlyPhe.....ValGluThrSer 420
1738 TCACAAACCTTTGACCCCTG..... 1755
:||||| :|||:||||| :|||:||||| :|||:|||||
421 GlyHisAspLeuSerIleGlyAspValIleValAspAlaLysGluTr 437
1756 .....ACTGTACGTCCTCGTCGT 1774
:||||| :|||:||||| :|||:||||| :|||:|||||
437 pLeuLeuAspProAspAspValSerIleGluThrLeuThrSerGlyArg 454
1775 CCAATCTACCTCCCTCCCAATTTACAGTGACTTCCAAAACAGACAGGAC 1824
:||||| :|||:||||| :|||:||||| :|||:|||||
454 snAsnThrGlyGluAsnGlnGlyTyThrThrGlyAspGlyThrLysGlu 470
1825 ACCAGCAAA.....TCCCCAGCCCTCTGTTATGCAAAATAT 1865
:||||| :|||:||||| :|||:||||| :|||:|||||
471 SerProLysGlyAsnSerIleSerLysProThrLeuThrAsnSerThrLe 487
1866 TCGCCAAGGAGCCTCCCAATTTCTAGGGCCAGTGTCACAGCCCTGATTG 1915
:||||| :|||:||||| :|||:||||| :|||:|||||
487 uGluGln.....IleLeuArgArgGlySerTyTrpValAsnIle 500
1915 AATCAGTGAATGGAANAACAGTTACCTTGAACATA...CTGGATAATGGA 1962
:||||| :|||:||||| :|||:||||| :|||:|||||
500 hrAlaAsnAsnArgIleTyTrpValAsnSerIleAsnLeuSerAsnGly 516
1963 GCAGGTGCTGATGCTACTAAGGATGACGGTGCTCTACTCAAGTATTTCAC 2012
:||||| :|||:||||| :|||:||||| :|||:|||||
517 SerLeuThrLeuHisThrLysArgAspGlyValLysIleAsnGlyAspI 533
2015 AACTTATGACAGCAATGTTAGATACAGTGTAAAAGTGGGGCTCTGGGAG 2062
|||:||||| :|||:||||| :|||:||||| :|||:|||||
533 eThrSerAsnGluAsnGlyAsnLeuThrIleLysAlaGlySerTrpValA 550
2063 GAGTTAAGCAGCAGCAGGAGAGTATACCCCGCAGAGTGGAGCAGCTG 2112
:||||| :|||:||||| :|||:||||| :|||:|||||
550 spValHis.....LysAsnIleThrLeuGlyThrGlyPheLeu 562
2113 TACATACCTGCTGGATTGAGAATGATGAATACAATGGAATCCACCAAG 2162
|||:||||| :|||:||||| :|||:||||| :|||:|||||
563 AsnIle.....ValAlaGlyAspSerValAlaPheGluArgGluG 576
2163 ACCTGAA..... 2169
:||||| :|||:||||| :|||:||||| :|||:|||||
576 yAspLysAlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrI 593
2170 .....ATTATAGGATGATGTTCAACACAG.....CAAGTGTGTTTC 2208
:||||| :|||:||||| :|||:||||| :|||:|||||
```


791 ACNAAAGCTCCAAACAGCAAAATCAAAAATGCAATCTCCGACACA 840
||:||||| |||||
577 snglulysasnAlaAsnLysGln..... 584
841 TGGGAAGTATCGGTGATTCGAGGACTTTAAGAAAACCACTCTATGAC 890
||:|||||ProLeuVa 587
585
891 AACACAGCCACCACAAATCCACCTCTCATTGCTGCAGATTGGACAAAGAA 940
:||||| :||| :|||
587 lthrAspSerThrAlaAlaThrValGlyAspLeu.....ArgLysL 601
941 TTGTGTGTTAGTCTTGCATTAATCTGGAAGCATGCGGACTGTAAACGC 990
: :||| :||| :|||
601 euGlyTrpValSerThrLysAsnGly.....ThrLysGluGlu 614
991 CTCATTCGACTGAATCAACGACGCCAGCTTTCTCTGCTGCACACATTGA 1040
|||:||||| :||| :|||
615 SerAsnGlnValLysGlnAlaAspGluVal..... 624
1041 GCTGGGTCTCGGTGGTGGATGGTGCATTTGCAGTGTGCCATGTAC 1090
||| :||| :||| :|||
625 .LeuPheThrGlyAlaGlyAlaAlaThrValThrSerLysSerGluAsnG 641
1091 AAAGTCAACTCATACAGATA.....AACAGTGC 1119
: :||| :||| :|||
641 lylshisThrIleThrValSerValAlaGluThrLysAlaAspSerGly 657
1120AGTGACGAGGACACACTCGCCAAAGATTACCT..... 1152
658 LeuGluLysAspGlyAspThrIleLysLeuLysValAlaAsnGlnAsnTh 674
1153GCAGCAGCTTCAGAGGAGCGTCCATCTGCAGCGGC 1189
: :||| :||| :||| :|||
674 rAspAsnValLeuThrValGlyAsnAsnGlyThrAlaValThrLysGly. 690
1190 TTCGATCGCATTTACTGTGATTAGGAAGAATAATCCAACTGATGATCT 1239
: :||| :||| :||| :|||
691GlyPheGluThrVal.....LysThrGlyAlaThrAspAlaAsp 703
1240 GAAATGTGCTGCTGCGGATGGGAAGACACACTATAAGTGGTGTCT 1289
: :||| :||| :||| :|||
704 Arg.....GlyLysValThrValLysAspAlaThrAl 714
1290 TAACAGGTCAACAAAGTGTGCCATCATCCACACAGTCTGTTGGGC 1339
||:||||| :||| :||| :|||
714 aAsnAspAlaAspLysLysValAlaThrValLysAspValAlaThrAlaI 731
1340 CCTCTCCAGCTCAAGACTAGAGGAGCTGCCAAATGACAGGAGTTTA 1389
: :||| :||| :||| :|||
731 leAsnSerAlaAlaThrPheValLysThrGluAsnLeuThrThrSerIle 747
1390 CAGACATATGCTTCAGATCAAGTTCAGACAATGGCTCATGATGCTTT 1439
: :||| :||| :||| :|||
748 Asp.....GluAspAsnProThrAspAsnGlyLysAspAsp..... 759
1440 TGGGGCCCTTTCATCAGGAAT.....GGAGCTGTCTCTCAGC 1477
||:||||| :||| :||| :|||
760AlaLeuLysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnL 775
1478 GCTCCATCCAGCTTGAGATGAGGATTAACCTC.....CAGAAC 1518
: :||| :||| :||| :|||
775 euLysValLysArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsn 791
1519 AGCAGTGTGAATGGCAGTGTGCGGACAGCAGCGTGGGAAGGA 1568
: :||| :||| :||| :|||
792 LeuGluValLysThrAlaLysValSerAspThrLeuThrIleGlyGlyAs 808
1569 CACTTTGTTTCTTATCACCTGGACACAGCAGCGCTCCCAAAATC..... 1611
:||| :||| :||| :|||
808 nthr.....ProThrGlyGlyThrThrAlaThrProLysValAsnIleT 823

1611
823 hrSerThrAlaAspGlyLeuAsnPheAlaLysGluThrAlaAspAlaSer 839
1612CTTCTCTGGATCC 1625
840 GlySerLysAsnValTyLeuLysGlyIleAlaThrThrLeuThrGluPr 856
1626 CAGTGGACAGCAAGAGTGGCTTTGTA.....GTGGACAAAACA 1666
||:||||| :||| :||| :|||
856 oSerAlaGlyAlaLysSerSerHisValAspLeuAsnValAspAlaThrL 873
1667 CCAAAATGGCTTACCTCCAAATCCAGGATTCGTAAAGTTGGCAGCTGG 1716
||| :||| :||| :|||
873 yLysSerAsnAlaAlaSerIleGluAspValLeuArgAlaGly..... 887
1716 AAATACAGTCTGCAAGCAAGCTCACAAACCTTCACCTGACTGTACGTC 1766
: :||| :||| :||| :|||
888 ...TrpAsnIleGlnGlyAsnGlyAsnValAspTyValAlaThrTy 903
1767 CCGTGGCTCCAATGTACC.....CTGCCTCCAATTACAG 1801
: :||| :||| :||| :|||
903 rAspThrValAsnPheThrAspAspSerThrGlyThrThrValThrV 920
1802 TGACTTCCAAACGACACAGGACACAGCAAAATCCCCAGCCCTCTGGTA 1851
||:||||| :||| :||| :|||
920 alThrGlnLysAlaAspGlyLysGly..... 928
1852 GTTATGCAAAATATTCGCAAGGAGCTCCCAATTCACGAGCCAGTGT 1901
: :||| :||| :||| :|||
929AlaAspValLysIleGlyAlaLys..... 936
1902 CACAGCCCTGATTGAATCAGTCAATGGAACACAGTTACC...TTGGAAC 1948
||:||||| :||| :||| :|||
937 .ThrSerValIleLysAspHisAsnGlyLysLeuPheThrGlyLysAspL 953
1949 TACTGATATGAGCAGGAGTGTGCTGCTACTAAGGATGACGCT..... 1992
|| :||| :||| :||| :|||
953 euLysAspAlaAsnAsnGlyAlaThrValSerGluAspAspGlyLysAsp 969
1993GTCTACTCAAGTATTTCACAACTTATCACACGAA 2027
: :||| :||| :||| :|||
970 ThrGlyThrGlyLeuValThrAlaLys.....ThrValIleAspAlaVa 984
2028 TGTAGATACAGTGTAAAAGTGGGCTCTGGGA..... 2061
: :||| :||| :||| :|||
984 lAsnLysSerGlyTrpArgValThrGlyGluGlyAlaThrAlaGluThrG 1001
2062GGAGTTAAGCGCAGCAGGAGAGTGTATCCCGCAGCAGT 2103
1004 lAlaThrAlaValAsnAlaGlyAsnAlaGluThrValThrSerGlyThr 1017
2104 GGAGCACTGTACATACCTGCTGGATTGAGAATGATGAAATACAATGGA 2153
: :||| :||| :||| :|||
1018 SerValAsnPhelysAsnGly.....As 1025
2154 TCACCAAGACCTGAAATTAATAAGGAT..... 2181
: :||| :||| :||| :|||
1025 nAlaThrThrAlaThrValSerLysAspAsnGlyAsnIleAsnValLysT 1042
2182 ..GATGTT.....CAACACAGCAAGTC 2202
: :||| :||| :||| :|||
1042 yrAspValAsnValGlyAspGlyLeuLysIleGlyAspAspLysLysIle 1058
2203 TGTCTTCAGCAGCAACATCC.....TCGGGAGCTCATTTGTG..... 2238
1059 valAlaAspThrThrThrLeuThrValThrGlyGlyLysValSerValPr 1075
2239GCTTCTGATGTCCAAATGCTCCCATACCTGATCTCTTCCAC 2281
: :||| :||| :||| :|||
1075 oAlaGlyAlaAsnSerValAsnAsn..... 1083
2282 CTGGCCAAATCACCGACCTGAGGCGGAAATTCACGGGGCAGCTCTCAAT 2331

1084AnLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsn 1098
2332 AATCTGACTGGACAGCTCTCTGGGGATGATATGACCATGGA 2373
1099 AsnLeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGly 1112

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT: AAR99393

seq_documentation_block:
ID AAR99393 standard; Protein; 2353 AA.

AC AAR99393;
DT 15-JAN-1997 (first entry)
XX Haemophilus adhesion protein HA2.
XX Haemophilus adhesion protein; HA2; hsf protein; vaccine.
XX Haemophilus influenzae type b strain C54.
OS WO9630519-A1.
XX 03-OCT-1996.
XX 22-MAR-1996; 96WO-US04031.
XX 24-MAR-1995; 95US-0409995.
XX (UWSL-) UNIV ST LOUIS.
XX (UNIW) UNIV WASHINGTON.
XX Barenkamp SJ, St Geme JW;
XX WPI; 1996-455364/45.
XX N-PSDB; AAT41476.

Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in
vaccines against H. influenzae infection.
Claim 5; Page 66-73; 120pp; English.
Haemophilus adhesion protein HA2 (AAR99393) is associated with the
formation of surface fibrils involved in adhesion to various host
cells; it is also referred to hsf (Haemophilus surface fibrils).
Its amino acid sequence was deduced from a genomic DNA clone
(AAT41476) derived from Haemophilus influenzae type b strain C65.
Large quantities of recombinant HA2 can be produced in transformed
prokaryotic or eukaryotic host cells, for use in vaccines against
H. influenzae infection.

Sequence 2353 AA;

alignment_scores:
Quality: 141.50 Length: 967
Ratio: 0.321 Gaps: 44
Percent Similarity: 45.605 Percent Identity: 18.718

alignment_block:

US-09-049-696-20 x AAR99393 ..
Align seg 1/1 to: AAR99393 from: 1 to: 2353
274 ACATGGAGACAAAGGCTGACTATGTCAGACCAAACTTGACACCTACAA 323
429 SerTriphysalalysAlaGlu..... 435
324 AATGCTGATGTTCTGTTCTGAGTCTACTCTCCAGGTAATGATGAAC 373
436AlaAsp.....ThrAspGlyAlaLeuGluGlyI 445

374 CCTACACTGACGAGATGGCACTCTGAGAGAAAGGTTGAAGGATCCAC 423
445 leSerLysAspGlnGluValLysAlaGlyGlu..... 455
424 CTCACCTCTGATTTTCATTCAGGAAAAAGTTAGCTGAATATGGACCACA 473
456ThrValThrPheLysAlaGlyLysAsnLeu...LysValLysGlnAs 470
474 AGGTAGGGCATTTGTCCATGAG.....TGGGCTC 502
470 pGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuThrGlyLeuThrs 487
503 ATCTACGATGGGAGTATTTGACGAGTACAATAATGATGAGAAATTCAC 552
487 erileThrLeuGlyGlyThrThrAsnGlyGlyAsnAspAlaLysThrVal 503
553 TATATCC...AATGGAGAAATACAGCAGTAAGATGTTTCAGAGGATTAC 599
504 IleAsnLysAspGlyLeuThrIleThrProAlaGlyAsnGlyGlyThrTh 520
600 TGTACAAATGTAGTAAAGATGTCAGGAGGAGGAGCTGTTACACCACAAA 649
520 rGlyThrAsnThrIleSerValThrLysAspGlyLysAlaGlyAsnL 537
650 GATGCACATTCATAAAGTAACAGGACTC.....TATCAAAAAGGATGT 693
537 ySalalileThrAsnValAlaSerGlyLeuArgAlaTyrAspAlaAsn 553
694 GAGTTTGTCTCCATCCCGCCAGAGGAGAGGCTTCTATATGTTTGC 743
554 PheAspValLeuAsnSerAlaThrAsp.....LeuAs 565
744 ACAACATGTTGATTCATAGTTGAATTCGTACAGAAACCAACCAACA 793
565 nArgHisValGluAspAlaTyrLysGlyLeuLeu.....AsnLeuAsnG 580
794 AAGAAGCTCCAAACAAAGCAAAATCAAAATTCGAATCTCCGAGCACATGG 843
580 lLysAsnAlaAsnLysGln..... 586
844 GAAGTGATCGTGATCTGAGGACTTTAAGAAACCACTCTCTATGACAAC 893
587ProLeuValTh 590
894 ACAGCCACCAAAATCCACCTTCTCATTCGTGAGATTTGGACAAAGAATTG 943
590 rAspSerThrAlaAlaThrValGlyAspLeu.....ArgLysLeuG 604
944 TGTGTTTGTCTTGCATAATCTGGAAGCATGGGAGCTGTACCCGCTC 993
604 lyTrrpValValSerThrLysAsnGly.....ThrLysGluGluSer 617
994 AATCGACTGATCAACAGCGCCAGCTTTTCTGCTGCAGACAGTTGAGCT 1043
613 AsnGlnValLysGlnAlaAspGluVal.....Le 627
1044 GGGGTCTCTGGGTTGGGATGTTGACATTTGACAGTGTGCCCATGTACAAA 1093
627 upheThrGlyAlaGlyAlaAlaThrValThrSerLysSerGluAsnGlyL 644
1094 GTGAACCTCATACAGATAAACAGTGGC..... 1119
644 ysHisThrIleThrValSerValAlaGluThrLysAlaAspCysGlyLeu 660
1120 ...AGTGACAGGACACACTCGCCAAAGATTACCT..... 1152
661 GluLysAspGlyAspThrIleLysLeuLysValAspAsnGlnAsnThrs 677
1153GCAGCAGCTTCAGGAGGAGCTCCATCTGCGAGCGGCTTC 1192
677 pAsnValLeuThrValGlyAsnAsnGlyThrAlaValThrLysGly.... 692
1193 GATCGGCATTTACTGTGATTAGGAAGAAATATCCAACCTGTGATCTCAA 1242

693GlypneGluThrVal.....LysThrGlyAlaThrAspAlaAspArg 706
1243 ATTGTGCTGTCGCGGATGGGAGACAAACACTATAAGTGGGTCTTAA 1292
707GlyLysValThrValLysAspAlaThrAlaAs 717
1293 CGAGGTCAAAACAAAGTGGTCCCATCATCCACAGCTCGCTTTGGGCCCT 1342
717 nAspAlaAspLysValAlaThrValLysAspValAlaThrAlaAla 734
1343 CTCGAGCTCAAGAACTAGAGGAGCTGTCCAAATAGCAGAGGTTTACAG 1392
734 snSerAlaAlaThrPheValLysThrGluAsnLeuThrSerIleAsp 750
1393 ACATATGCTTCAGATCAAGTTCAGAAATGGCTCATTCATGCTTTGG 1442
751GluAspAsnProThrAspAsnGlyLysAspAsp..... 761
1443 GGCCCTTTCATCAGGAAT.....GGAGCTGCTCTCAGCGCT 1480
762 .AlaLeuLysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeu 778
1481 CCATCCAGCTTGAGAGTAAGGATTAAACCTC.....CAGAACAGC 1521
778 ysValLysArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeu 794
1522 CAGTGGATGAATGCCAGAGTGTGGAGCAGCACCGCTGGGAAGACAC 1571
795 GluValLysThrAlaLysValSerAspThrLeuThrIleGlyGlyAsnTh 811
1572 TTTGTTTCTTATCACCTGGACAGCAGCTCCCAATC..... 1611
811 r.....ProThrGlyGlyThrAlaThrProLysValAsnIleThrS 826
1611 1611
826 erThrAlaAspGlyLeuAsnPheAlaLysGluThrAlaAspAlaSerGly 842
1612CTTCTCTGGGATCCCG 1628
843 SerLysAsnValTyrLeuLysGlyIleAlaThrThrLeuThrGluProSe 859
1629 TGCAGAGAACAAGTGGCTTTCTGA.....GTGGCAAAAACACCA 1669
859 rAlaGlyAlaLysSerSerHisValAspLeuAsnValAspAlaThrLysL 876
1670 AAATGGCTACCTCCAAATCCAGGATGCTTAAGTGTGGCACTTGGAAA 1719
876 ysSerAsnAlaAlaSerIleGluAspValLeuArgAlaGly..... 889
1720 TACAGTCTGAACAGCTCACAACCTTGACCTGACTGCTACGTCGCCG 1769
890 TrpAsnIleGlnGlyAsnGlyAsnAsnValAspTyrValAlaThrTyrAs 906
1770 TGGCTCCAATGCTACC.....CTGCCTCCCAATTACACTGA 1804
906 pThrValAsnPheThrAspAspSerThrGlyThrThrThrValThrVal 923
1805 CTTCCAAAACGAACAGGACACCAAGCAAAATTCGCCGCCCTCTGGTAGT 1854
923 hrGlnLysAlaAspGlyLysGly..... 930
1855 TATGCMAATATTCGCNAGAGAGCTCCCAATTCCTCAGGCCAGTGTAC 1904
931 ...AlaAspValLysIleGlyAlaLys.....Th 939
1905 AGCCCTGATTGAATCACTGAATGGAAGAAACAGTTACC.....TTGGAAC 1951
939 rSerValIleLysAsnGlyLysLeuPheThrGlyLysAspLeu 956
1952 TGGATAATGAGCAGGTGCTGTGATGCTACTAAGGATGACGGT..... 1992
||| ... ||||| :||||| :|||||

ysAspAlaAsnAsnGlyAlaThrValSerGluAspAspGlyLysAspThr 972
1993GTCTACTCAAGTATTTTCACTATATGACACGATGG 2030
973 GlyThrGlyLeuThrAlaLys.....ThrValIleAspAlaValAs 987
2031 TAGATACAGTGTAAAGTCCGGCTCTGGGA..... 2061
987 nLysSerGlyTyrArgValThrGlyGluGlyAlaThrAlaGluThrGly 1004
2062GGAGTTTAACGACGACGAGAGTGTATCCCGCAGCAGTGGA 2106
1004 lathrAlaValAsnAlaGlyAsnAlaGluThrValThrSerGlyThrSer 1020
2107 GCACCTGTACATACCTGGCTGGATTGAGAATGATGAATAACAATGGAATCC 2156
1021 ValAsnPheLysAsnGly.....AsnAl 1028
2157 ACCAAGACCTGAATTAATAAGGAT.....G 2182
1029 aThrThrAlaThrValSerLysAspAsnGlyAsnIleAsnValLysTyrA 1045
2183 ATGTT.....CAACACAAGCAAGTGTGT 2205
1045 spValAsnValGlyAspGlyLeuLysIleGlyAspAspLysLysIleVal 1061
2206 TTCAGCAGAACATCC.....TCGGAGGCTCATTTGTG..... 2238
1062 AlaAspThrThrThrLeuThrValThrGlyGlyLysValSerValProAl 1078
2239GCTTCTGATCTCCCAATGCTCCCATACCTGATCTCTCCACCTG 2284
1078 aGlyAlaAsnSerValAsnAsn..... 1085
2285 GCCAAATCACCGACCTGAAGCGGAAATTCACGGGCGAGCTCATTAAT 2334
1086AsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsnAsn 1101
2335 CTGACTTGGACAGCTCTCTGGGATGATTATGACCATGGAAACAGCTCACAA 2384
1102 LeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyGluSerGluGl 1118
2385 GTATATC..... 2391
1118 yGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysAlaG 1135
2392ATTGGAATAAGTACAAAGTATTTCTGATCTCAGACAGCAAGTTC 2433
1135 yLysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSerLeu 1151
2434 ATGAATCTCTCAAGTGAATACTACTGCTCTCATCCCAAGGAAGCAAA 2483
1152 GluAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAlaAs 1168
2484 C.....TCGTAGAGAACTCTTTTGTGTTAAACCAAGAAACATTTACTTTT 2526
1163 nGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrIleT 1185
2527GAAATGGCAGAGAT 2541
1185 hrLeuAlaAsnGlyAlaAlaAlaGlyThrAspAlaSerAsnGlyAsnThr 1201
2542 CTTTTCATGCTATTTCAGGCTGTGTGATAA..GGTCGATCTGAAATCAGAAA 2590
1202 IleSerValThrLysAspGlyIleSerAlaGlyAsnLysGluIleThrAs 1218
2591 TATCCAACATTTGCACGAGTATCTTTGTTTATTTCTCCACAGACTCCGCCA 2640
1218 nValLys.....SerAlaLeuLysThrTyrLysAspThrGlnAsnThrA 1233
2641 GAGACACCTAGTCTCTGATGAACAGCTCTGCTCTTGTCTCTTAATATCATAT 2690
1233 laAspGlu..... 1235


```
335 lySerAspIleAlaLysThrGlyGlyPhe.....ValGluThrSer 348
1738 TCACAAACCTGGACCCCTG.....
349 GlyHisAspLeuSerIleGlyAspValIleValAspAlaLysGluTr 365
1756 .....ACTGTCAGTCCGCTGCT 1774
365 pLeuLeuAspProAspAspValSerIleGluThrLeuThrSerGlyArgA 382
1775 CCAATGCTACCTGCTCCAAATACAGTACCTCCAAACGACACAGGAC 1824
382 snAsnThrGlyGluAsnGlnGlyThrThrGlyAspGlyThrLysGlu 398
1825 ACCAGCAAA.....TTCCCCAGCCCTCTGGTAGTTTATGCAAAATAT 1865
399 SerProLysGlyAsnSerIleSerLysProThrLeuThrAsnSerThrLe 415
1866 TCGCCNAGGAGCTCCCAATCTCAGGCCAGTGTACAGCCCTGATTTG 1915
415 uGluGln.....IleLeuArgArgGlySerTyrrValAsnIleT 428
1916 AATCAGTGAATGGAACAAACAGTTACCTTGGAACTA...CTGGATATGGA 1962
428 hrAlaAsnAsnArgIleTyrrValAsnSerSerIleAsnLeuSerAsnGly 444
1963 GCAGGTGCTGTACTACTAAGTATGACGGTGTCTACTCAAGTATTTCAC 2012
445 SerLeuThrLeuHisThrLysArgGlyValLysIleAsnGlyAspIle 461
2013 ACTTATGACAGCAATGTTAGTACAGTAAAGTGGGGCTCTGGGAG 2062
461 eThrSerAsnGluAsnGlyAsnLeuThrIleLysAlaGlySerTrpValA 478
2063 GAGTTAAACGACGACGAGAGTGTATATACCCAGCAGAGTGGACACTG 2112
478 spValHis.....LysAsnIleThrLeuGlyThrGlyPheLeu 490
2113 TACATACCTGGTGGTATTGAGATGATGAATACAAATGGAATCCACCAAG 2162
491 AsnIle.....ValAlaGlyAspSerValAlaPheGluArgGluG 504
2163 ACCTGNA..... 2169
504 yAspLysAlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrI 521
2170 .....ATTAATAGGATGATGTTCAACACAAG.....CAAGTGTGTTTC 2208
521 leThrValAsnLysAspAspLysGlnPheArgPheAsnAsnValSerIle 537
2209 AGCAGAACATCTCTGGGAGGCTCATTTGTGGCTTCGTATGTCGCCAAATGC 2258
538 AsnGlyThrGlyLysGlyLeuLysPheIleAlaAsnGln..... 550
2259 TCCCATACCTGATCTCTCCACCTGGCCAAATCACC.....GACC 2299
551 .....AsnAsnPheThrHisLysPheAspG 559
2300 TGAAGCGGAAATTCACGGGGGAGTCTCATTAATCTGACT..... 2340
559 lGluIleAsnIleSerGlyIleValThrIleAsnGlnThrThrLysLys 575
2341 .....TGGACAGCTCTGGGGGATGATTATGACCAT..... 2370
576 AspValLysTyrrTrpAsnAlaSerLysAspSerTyrrTrpAsnValSerSe 592
2371 .....GGAACAGCTCACAAGTAT.....ATCATTCGAA 2398
592 rLeuThrLeuAsnThrValGlnLysPheThrPheIleLysPheValAspS 609
2399 TAAGTACAGTATCTTGTATCTCAGAGCAAGTTCATCAATCAATCTCTCAA 2448
.....
```

```
609 erGlySerAsnSerGlnAspLeuArgSerSerArgArgSerPheAlaGly 625
2449 GTCAATACTACTGCTCTCATCCCAAGGAA.....GCCAA 2483
625 ValHisPheAsnGlyIleGlyGlyLysThrAsnPheAsnIleGlyAlaAs 642
2483 CTCTGAGGAGTCTTTTCTTTAAACCAGAAAACATTACT 2523
642 nAlaLysAlaLeuPheLysLeuLysProAsnAlaAlaThr 655
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT:AA41725

seq_documentation_block:

ID AA41725 standard; Protein: 1536 AA.

XX AA41725;

DT 26-APR-1994 (first entry)

XX High molecular weight protein 1 (HMW1).

XX HMW; high molecular weight protein; virus; vaccine; influenza;

XX epitope; immunity; haemophilus influenzae; gene cluster.

XX Haemophilus influenzae.

XX Key Location/Qualifiers

FT Misc-difference 668..677

FT /note= "Possibly incorrect sequence. Alternative

sequence for this region is LNVSESGEFN.

(See comments)."

XX WO3319090-A.

XX 30-SEP-1993.

XX 16-MAR-1993; 93WO-US02166.

XX 16-MAR-1992; 92GB-0005704.

XX (BARE/) BARENKAMP S J.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Barenkamp SJ;

XX WPI; 1993-320683/40.

XX N-PSDB; AAQ49508.

XX High molecular weight surface proteins - of non-typeable

haemophilus which exhibit immunogenic properties

XX Cl3im 3; Figure 2/10; 100pp; English.

XX The isolation and purification of the high molecular weight protein enables the identification of the major protective epitopes of the protein by conventional epitope mapping. These epitopes can then be synthesised using standard techniques and incorporated into fully synthetic or recombinant vaccines. This sequence is claimed to be the same as that given in AA41723 (High molecular weight protein 1) although it does, differ slightly. (Repeated regions which are possibly incorrect and occur in the corresponding nucleotide coding sequence contribute to these differences).

XX Sequence 1536 AA;

XX alignment_scores:

XX Quality: 139.50 Length: 938

XX Ratio: 0.319 Gaps: 49

XX Percent Similarity: 46.588 Percent Identity: 20.256

XX alignment_block:

XX US-09-049-696-20.x AA41725

Align seg 1/1 to: AAR41725 from: 1 to: 1536

```
217 TTTGAAGCTACA.....GGAAGCGATTTTATTTCATAAATGT 254
|||||
631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnII 643
|||||
255 TGCATTTTGTCTCTGACATGGAACAGGCTGACTATGTGAGAC 304
|||||
643 eSerMetValLeuPro.....LysAsnGluSerGlyTyrAspLysp 657
|||||
305 CAATACTTGAGACCTACAAAAT..... 327
|||||
657 heLysGlyArgThrTyrTrpAsnLeuThrSerLysValAspMetIleAsn 673
|||||
328 ...GCTGATGTCGGTGTGCTGACTACTCTCCCTCAGGTAATGAT..... 369
|||||
674 SerLysAspAlaLeuThrIleAspSerArg.....GlySerAspSerAl 688
|||||
370 .....GAACCTACACT.....G 382
|||||
688 aGlyThrLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysA 705
|||||
383 AGCAGATGGCACTGTGGAGAGA.....AGGTGAAGATCCACCTC 426
|||||
705 spThrThrPheAsn.ValGluArgAsnAlaArgValAsnPheAspIleLy 721
|||||
427 ACTCCTGATTCATTCGAGGAGAAAGTTAGCTGAATGTCACCACAAGG 476
|||||
721 salaProfileGlyIleAsnLysTyrSerSerLeuAsnTyrAlaSerPheA 738
|||||
477 TAGGGCATTTGCCATGAGTGGGCTCATCTACGATGGGGAGTATTGACG 526
|||||
738 snGlyAsnIleSerValSerGly.....GlyGlySerValAsp 750
|||||
527 AGTACAATTAATGATGAGAAATCTACTATCAATGGAGATACA..... 572
|||||
751 PheThrLeuLeuAlaSerSerSerAsn.ValGlnThrProGlyValValI 767
|||||
573 ..AGCAGTAAGATGTCACAGGATTTACTGGTACAAATGTAGTAAAGAA 620
|||||
767 leAsnSerLysTyrPheAsn.ValSerThrGlySerSerLeuArgPheLy 783
|||||
621 GTGTCAAGGA.....GGCAGCTGTGTACACCAAAAGATCCACAT 658
|||||
783 sThrSerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrL 800
|||||
659 TCAAT.....AAAGTAACAGGACTCTATGAAAAGGATGTGAG 696
|||||
800 euAsnAlaThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAsp 816
|||||
697 TTTGTTCTCAATCCCGCAGAGGAGAGGCTTCTATAATGTTT..... 741
|||||
817 GlyMetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluG 833
|||||
742 .....GCACAACATGTTGATTCTA 760
|||||
833 yGlyLysMetArgPheGlySerArgLysAlaValThrGluIleGluGlyA 850
|||||
761 TAGTTGAATTTCTGACAGAA..... 780
|||||
850 snValThrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPhe 866
|||||
781 CAATACCAACAAGAGCTCCAAACAGCAAAAT.....CAAAA 821
|||||
867 AspAsnHisGlnLysProLeuThrIleLysLysAspValIleIleAsnSe 883
|||||
822 ATGCAATCTCGAAGACATGGAAGTG..... 849
|||||
883 rGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeu 900
|||||
850 ..ATCCGTGATTCGAGGACTTTAAGAAACCACTCTCTATGACA..... 891
|||||
```


Align seq 1/1 to: AAW56322 from: 1 to: 2039

871 AAGAAACCACTCCTATGACAACACGCCAACAAATCCC..... 909
||| :|||:||||:||||:||||: ||| :|||
1179 LysAlaAsnSerProIleThrValGluProSerThrAspAsnAsnLysLy 1195
910ACCTTCCTCATCTGCTCCAGATTGGACAAAGAATTCTGTGT 949
||| :|||:||||:||||: ||| :|||
11195 sLysThrPheThrValGlyLeuMetLysAspIleGluGlyValAsnSerI 1212
950 TAGTCCCTTGACAAATCTGGA.....AGCATGGCGACTGGTAACCGC 990
||| :|||:||||:||||: ||| :|||
1212 IeThrPheAspLysSerGlyGlnAspLeuAsnGlnValThrGly..... 1226
991 CTCAAATCGACTGAATCAACGAGCCAGCTTTTCCGTGCTGCAGACAGATGA 1040
||| :|||:||||:||||: ||| :|||
1227ArgMetSerSerAlaGlyLeuThrPheLysLysGlyAspThrTh 1241
1041 GCTGGGTCTGGGTGGGATG.....GTGACATTTGCACAGT 1078
1241 rAsnGlySerThrThrPheAlaGluAspGlyLeuThrIleAspSerT 1258
1079 CTGCCCAT...GTACAAGTGAATCATACAGATAAACC..... 1113
1258 hrTrAsnSerAlaGlnThrAsnLeuValLysValSerArgaspGlyPhe 1274
1114ACTGGCAGTGCACGGGACACACTC..... 1137
1275 SerValLysAsnGlySeraspGluSerLysLeuAlaSerThrLysLeuSe 1291
1138GCCAAAAGATTACTGCAGCAGCTTCCAGNA..... 1167
1291 rIleGlyAlaGluAsnAlaGluHisValGluValThrLysSerGlyIleA 1308
1167 1167
1308 IaLeuLysAlaAspAsnThrSerAspLysSerSerIleThrLeuAlaGln 1324
1168GGCAGCTCCATC....TGCAG 1184
1325 AspAlaIleThrLeuAlaGlyAsnAlaThrGlyThrAlaIleLysLeuthr 1341
1185 CGGGCTTCGATCGGCTATTACTGTGATTAGGAGAATAATCCAACTGATG 1234
||| :|||:||||:||||: ||| :|||
1341 rGlyValAlaAspGlyAsnIleThrValAsnSerLysAspAlaValAsnG 1358
1235 GATCTGAATT.....GTGCTGCTGCGGATGGG..... 1263
1358 lyGlyGlnLeuarghrThrLeuLeuGlyValAspSerGlyAlaLysIleGly 1374
1264GAAGACAACACTAATAAGTGGGTGCTTTAACGAGTCAAACAAAG 1307
||| :|||:||||:||||: ||| :|||
1375 GlyThrGluLysThrThrIleSerGluAlaIleSerAspValLysGlnAl 1391
1308 TGGTGCCCATCCACACAGTGGCTTTGGGGCCCTCTGCAGCTCAAGAAC 1357
||| :|||:||||:||||: ||| :|||
1391 a.....LeuThrAspAlaThrLeuAlaTyrlLysAlaAspAsnLysA 1405
1358 TAGANGNGCTGCCAAATATGACGAGGTTCACAG.....ACA 1395
||| :|||:||||:||||: ||| :|||
1405 snGlyLysThrValLysLeuThrAspGlyLeuAsnPheThrSerThrThr 1421
1396 TATGCTTTCAGACTCAAGTTCACAGCAATGGCCTCAATTGAT..... 1434
||| :|||:||||:||||: ||| :|||
1422 AsnIleasPalaserValGluAspAsnGlyValLysPheThrLeuLy 1438
1435GCITTTGGGGCCCTTTCATCAG 1456
||| :|||:||||:||||: ||| :|||
1438 sasPLysLeuThrGlyLeuLysThrIleAlaThrGluSerLeuAsnAlas 1455
1457 GAATGAGAGCTGCTCTCAGCGCTCCATCCAGCTT.....GAGAGTAAAG 1500

1455 erGlnAsnIleLeAlaGlyGlyThrValThrValGlyGlyGluThrGlu 1471

1501 GGATTAAACCTCCAGAACACGCCAGTGGATCAAT 1533
|||||:|||||:||||| |||
1472 GlyIleValIeuThrLysSerGlySerGlyAsnAspArgThrLeuSerLe 1488
|||:|||||:||||| |||
1533 1533

1488 userGlyAlaGlyAsnAlaAlaThrAspGlyTleLysValSerGlyVal 1505

1534 GGCACAGTGATCGTCAGCAGCACCGCTGGGAAG GAC 1569
|||||:|||||:|||||:|||||:||||| |||
1505 ysAlaGlyThrAlaAspThrAspAla...ValasNLysGlyGlnLeuAsp 1520
|||||:|||||:|||||:|||||:||||| |||
1570 ACTTTGTTCCTT ATCACCTG 1589
||||| |||:|||||

1521 LysLeuPheLysAlaIleAsnAspAlaLeuGlyThrThrAspLeuAlaVa 1537

1590 GACNACGACGCTCCCACAATCTTCCTCGGATCCC...AGTGGACAGA 1636
|||:||||| |||:|||||:|||||:|||||:||||| |||
1537 lThrLysAsnProAsnGlnThrSerIlePheAsnProIleAsnGlyThrA 1554
|||||:|||||:|||||:|||||:||||| |||
1637 AGCAAGCTGGCTTT GTACTGGCAAAAAACC 1668
|||||:|||||:|||||:|||||:||||| |||

1554 laProThrThrPheLysAspAlaValAspLysLeuThrThrAlaValAsn 1570

1669 AAATGCCCTACTCTCAA ATCCAGGCAAT 1697
|||||:|||||:||||| |||

1571 ThrGlyTrpGlySerLysValIleGlyIleLeuAlaThrGlyIleAspGlyIl 1587

1693 TGCTAAGCTTGCCTCTGGAATACAGTCTCCAAGCAAGC.....T 1738
|||:|||||:||||| |||:|||||:||||| |||

1597 e...AspAlaGlyAsnLysIleSerAsnValAlaAspGlyAspIleS 1603

1739 CACAAAACCTTGACCTGACTGTACGTCCTCCGT 1770
||| |||:|||||:||||| |||

1603 erProThr-SerGlyAspValValThrGlyArgGlnLeuTyrrAlaLeuMet 1619

1771GGTCCAATCC 1781

1620 GlnLysGlyIleArgValTrpGlyAspGluValSerProThrLysThrGl 1636

1782 TACCCTGCCTCCAAATTACAGTGACTCCAAACGACAGACACCACCA 1831
||| |||:|||||:||||| |||:|||||:||||| |||

1636 nThrThrAlaProThrAlaSerSerThrGlnGlyGlyAlaThrThrAlaA 1653

1832 AATTCCCCAGCCTCTGGTAGTTTATGCAAAATATTCGCAAGGA 1875
::: |||:|||||:|||||:|||||:||||| |||

1653 snThrAlaGlyGlyValAlaProAlaGlyAsnValalathrcgylaspile 1669

1876GCCTCCCAATCTCAGGCCAGTGTACGCCCTGATT 1914
|||:|||||:||||| |||:|||||:||||| |||

1670 AlaProThrGlnProAlaLeuProGluMetLysThrAlaLeuValclysA 1686

1915 GAAT 1918

1685 phiLSLeuAlaValProLeuGlyGlySerLeuLysIleHisGlyAspHisA 1703

1919 CAGTGAATGGAAACACAGTTACCTTGGAACTACTG 1953
:|||:|||||:|||||:|||||:||||| |||:|||||:||||| |||

1703 snValLysThrThrIleSerAlaGlyAsnGlnValGlyIleSerGluIn 1719

1954 :GATAAT 1959

1720 ProAsnIleSerIleGluAsnAsnLeuValIleGlySerAsnLysProGl 1736

1960 GGACAGGTGGCTGATGCTACTA 1981

1736 ulysAlaLysLeuAlaAlaGlnGluGlyAsnAlaLeuValIleThrAsnL 1753

1982 AGGATGACGGTGCTACTCAAGGTATTTCACACTTATGACACGAATGGT 2031
|||||:|||||:|||||:|||||:||||| |||:|||||:||||| |||

1753AsnAsn 1764

```
2032 AGATACAGTCTAAAGTGGGGCTCTGGGAGGAGTAAACCCAGACG 2081
    : : : : : : : : : : : : : : : : : : : : : : : :
1765 GluLysAsnMetLeuVal.....LeuSerAspLysAlaLysProAr 1779
    : : : : : : : : : : : : : : : : : : : : : : : :
2082 GAGAGTGATACCCAGCAGTGGAGCAGTGTACATACCTGGCTGGATTG 2131
    : : : : : : : : : : : : : : : : : : : : : : : :
1779 galavalleuaspGlyGlnAsnGlyAlaLeuThrLeuValGly..... 1793
    : : : : : : : : : : : : : : : : : : : : : : : :
2132 AGAATGATGAATAACAA.....TGAATCCACCAAGACCTGAAT 2172
    : : : : : : : : : : : : : : : : : : : : : : : :
1794 ..AsnAspAspSerGlnValThrLeuSerSerLysLysGlyLysAsp 1809
    : : : : : : : : : : : : : : : : : : : : : : : :
2173 AATAAGCATGATGTTCAACAGCAAGTGTGTTTCAGCAGACATCCCTC 2222
    : : : : : : : : : : : : : : : : : : : : : : : :
1810 AspGlyAsnAspLeuSerArgLeuSerValThrThrGluArgThrAsnAl 1826
    : : : : : : : : : : : : : : : : : : : : : : : :
2223 GGAGGCTCATTTGTCCTCTGATGTCCTCCAAATGCTCCCATACCTGATC 2272
    : : : : : : : : : : : : : : : : : : : : : : : :
1826 aspGlyGlnLeuGluLysValGluThrSerPheAlaThrMetAspAsp. 1842
    : : : : : : : : : : : : : : : : : : : : : : : :
2273 TCTTCCCACTGGCCCAATACCGACCTGAAGCGGGAATTCACGGGGC 2322
    : : : : : : : : : : : : : : : : : : : : : : : :
1843 .....GlyLeuLysPheLysAlaAspGlyAsp 1851
    : : : : : : : : : : : : : : : : : : : : : : : :
2323 AGTCTCATTAATCTG.....ACTGGACAGCTCTCTGGGATGA 2360
    : : : : : : : : : : : : : : : : : : : : : : : :
1852 LysValIleAsnLysLysLysLeuAsnGluThrValGluIleValGlyAsp 1868
    : : : : : : : : : : : : : : : : : : : : : : : :
2361 TTATGACCATGGAACAGCTCACAAAGTATATCATTCGAATAAGTACAGTA 2410
    : : : : : : : : : : : : : : : : : : : : : : : :
1868 .....AsnValThrThrSerI 1874
    : : : : : : : : : : : : : : : : : : : : : : : :
2411 TTCTTGAT.....CTCAGACAGCAAGTCAATGAATCTCTCAAGTG 2451
    : : : : : : : : : : : : : : : : : : : : : : : :
1874 leThrAspAsnLysValLysValSerLeuAsnLysLysIleAlaIle 1890
    : : : : : : : : : : : : : : : : : : : : : : : :
2452 AATACTACTGCTCTCATCCCAAGGAGCAACTCTGAGAGTCTTTT 2501
    : : : : : : : : : : : : : : : : : : : : : : : :
1891 AspGluValIleProAsnThrAspProAspAlaGln..... 1903
    : : : : : : : : : : : : : : : : : : : : : : : :
2502 GTTTAAACAGAAAACATTACTTTTCAAAATGGCACAGATCTTTTCATTTG 2551
    : : : : : : : : : : : : : : : : : : : : : : : :
1904 ....LysGlyAspSerIleValIleAsnAsnGly.....G 1914
    : : : : : : : : : : : : : : : : : : : : : : : :
2552 CTATTCAGGCTGTGTAAGGTC 2574
    : : : : : : : : : : : : : : : : : : : : : : : :
1914 lylleHisAlaGlyAsnLysVal 1921
    : : : : : : : : : : : : : : : : : : : : : : : :
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT: AAB01847

seq_documentation_block:

ID AAB01847 standard; Protein: 1095 AA.

XX AC AAB01847;

XX DT 11-SEP-2000 (first entry)

XX DE Haemophilus influenzae strain 12 mature HMW1A protein, SEQ ID NO:69.

XX KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.

XX OS Haemophilus influenzae strain 12.

XX PN WO200020609-A2.

XX PD 13-APR-2000.

XX PF 07-OCT-1999; 99WO-CA00938.

```
XX 07-OCT-1998; 98US-0167568.  
PR 08-DEC-1998; 98US-0206942.  
XX (CONN-) CONNAUGHT LAB LTD.  
XX Loosmore SM, Yang Y, Klein MH;  
XX WPI; 2000-303789/26.  
DR N-ESDB; AAA52196.
```

XX Nucleic acid-molecule for producing recombinant high molecular weight
XX proteins of Haemophilus which are used as a vaccine to provide
XX protection against Haemophilus induced diseases in humans -
XX Claim 8; Fig 28F-Q; 307pp; English.

The invention relates to the recombinant production of Haemophilus
influenzae high molecular weight (HMW) proteins in Escherichia coli. The
expression construct used to effect recombinant expression comprises a
promoter functional in E. coli (e.g., the T7 promoter) operably linked
to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA,
hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
and the hmwB and hmwC genes encode accessory proteins which are
responsible for post-translational processing and secretion of the HMW
proteins. The modified hmwABC operon used in the expression construct of
the invention contains an A gene modified such that it encodes only the
mature HMW. The invention also discloses hmwA genes (AAA52175-A52198)
and HMW proteins (AAB01824-B01849) from the non-typeable H. influenzae
strains Joyce, KI, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
vectors are used for the production of recombinant H. influenzae HMW
proteins which can be used as vaccines to mediate a humoral or
cell-mediated immune response to provide protection against diseases in
humans caused by H. influenzae (e.g., otitis media, epiglottitis,
pneumonia and tracheobronchitis). The HMW proteins are also useful as
antigens in immunoassays for detecting antibodies against Haemophilus,
HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
HMW proteins can be used to isolate and clone hmw genes from other
non-typeable strains of Haemophilus via hybridisation reactions. The
present sequence represents a mature HMW protein from a non-typeable
strain of H. influenzae.

XX Sequence 1095 AA;

alignment_scores:	Quality: 138.50	Length: 936
	Ratio: 0.318	Gaps: 49
Percent Similarity:	46.581	Percent Identity: 20.406

alignment_block:

US-09-049-696-20 x AAB01847

Align seg 1/1 to: AAB01847 from: 1 to: 1095

```
217 TTGAGAGCTACA.....GGAAAGCGATTTTATTTCAAAATGT 254
    : : : : : : : : : : : : : : : : : : : : : : : :
190 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnI 202
    : : : : : : : : : : : : : : : : : : : : : : : :
255 TGCCATTTTCATTCCTCGAACATGGGAAGACAAGCGTGACTATGTGAGAC 304
    : : : : : : : : : : : : : : : : : : : : : : : :
202 eSerMetValLeuPro.....LysAsnGluSerGlyTyrAspLysP 216
    : : : : : : : : : : : : : : : : : : : : : : : :
305 CAAACTTCAGACCTACAAAATGCTGCTCTCTG...GTTGCTGAGTCT 351
    : : : : : : : : : : : : : : : : : : : : : : : :
216 helysGlyArgThrTyrrPasnLeuThrSerLeuAsnValSerGluSer 232
    : : : : : : : : : : : : : : : : : : : : : : : :
352 .....ACTCTCCAGGTAATGAT..... 369
    : : : : : : : : : : : : : : : : : : : : : : : :
233 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspAlaGlyTh 249
    : : : : : : : : : : : : : : : : : : : : : : : :
```

```

370 .....GAACCTACACT.....GAGCAGA 388
249 rLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThrT 266
389 TGGCAACTGTGGAGAGA.....AGGGTAAAGGATCCACCTCACVCT 432
266 hrPheAsn.ValGluArgAsnAlaArgValAsnPheAspIleLysAlaPr 282
433 GATTTCATTCAGCAAAAAAGTTAGCTGAATATGGACCACCAAGGTAGGC 482
282 oileGlyIleAsnLysTyrSerSerLeuAsnTyrAlaSerPheAsnGlyA 299
483 ATTGTCTCATGAGTGGCTCATCTACGATGGGAGTATTTCACGAGTACA 532
299 snIleSerValSerGly.....GlyGlySerValAspPheThr 311
533 ATAATGATGAGAAATTCATTATTCCAATGGAAGAATACA.....AGCA 576
312 LeuLeuAlaSerSerSerAsn.ValGlnThrProGlyValValIleAsnS 328
577 GTAACATGTTCCAGCAGGATATTACTGGTACAAATGTAGTAAAGAAGTGTCA 626
328 erLysTyrPheAsn.ValSerThrGlySerSerLeuArgPheLysThrSe 344
627 GGGA.....GGCAGCTGTTACACCAAAAGATGCACATTCAT. 663
344 rGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeuAsnA 361
664 .....AAGTACAGGACTCTATGAAAGAGGATGAGCTTTGCT 702
361 laThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspGlyMet 377
703 CTCCAATCCGCCACGAGGAGGAGCTTCTATAATGTTT..... 741
378 ileGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyGlyAs 394
742 .....GCACACATGTTGATTCTATAGTTG 766
394 nileThrPheGlySerArgLysAlaValThrGluIleGluGlyAsnValT 411
767 AATCTGTACAGAA.....CAAAAC 786
411 hrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPheAspAsn 427
787 CACAACAAAGAGCTCCAAACAAAGCAAAAT.....CAAAATGCAA 827
428 HisGlnLysProLeuThrIleLysLysAspValIleAsnSerGlyAs 444
828 TCTCCGAACCATGGGAGTG.....ATCC 853
444 nLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuThrValG 461
854 GTGATTCCTGAGGACTTTAAGAAACACCTCCATGACA..... 891
461 luSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsnValGly 477
892 .....ACACAGCCCAATCCACCTTCTCATCTCCTCCAGATTGG 932
478 GlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLysGlyGl 494
933 ACNAGAATTGTGTTTGTAGTCTTCGACAAATCTGGAAGCATGGGACTG 982
494 YAlaArgPheLysAspIleAspAsnSerLysAsnLeuSerIleThrA 511
983 GTAACGCCCTC.....AATGACTGAATCAA 1008
511 snSerSerThrTyrArgThrIleIleSerGlyAsnIleThrAsnLys 527
1009 GCAGCCAGCTTTTCTGCTGGCAGACAGTTGAGCTGGGCTCTGGGTTGG 1058
528 AsnGlyAspLeuAsnIleThrAsnGlu.....GlySer..... 538
1059 GATGGTGACATTTGACAGTGCTGCCCATGTACAAAGTGAACCTATACAG. 1107

```

```

539 .....AspThrGluMetGlnIleGlyAspValSerGlnL 551
1103 .....ATAAACAGTGGCACTGACAGGACACACATGCCCAAAAAGA 1146
551 ysGluGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrLysGln 567
1147 .....TTACCTGCAGCAGCTTCAGGAGGAGCTCCATCTCGACGGGCT 1190
568 IleThrIleLysAlaGlyValAspGlyGluAsnSerSerAspAlaIleTh 584
1191 TCGATCGGCATTTCATGCTAGTAGGAAGAAA..... 1221
584 rAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnA 601
1222 .....TATCCAACCTGATGGA 1236
601 splLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGly 617
1237 TCTGAATTTGCTGCTGACGATGGGGAAGACACACATATAAGTGGG... 1284
618 SerAsp.....LeuThrIleGlyAsnThrAsnSerAlaAspGlyTh 631
1285 .....TGCTTTAACGAGGTCAACAAAGT.....GGTG 1312
631 rAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 648
1313 CCATCATCCACACAGTCGCTTTGGGCCCTCTGCACCTCAAGAACTAGAG 1362
648 laAspGlyHisLysValThrLeuHis..... 656
1363 GAGCTGTCCAAATG...ACAGGAGCTTTACAGACATATGCTTCAGATCA 1409
657 .....SerLysValGluThrSerGlySerAsnAsnThrGluAspSe 671
1410 AGTTACAGACATGGCTCATTTGTTGGGCCCTTTTCATCAGGAA 1459
671 rSerAspAsnAsnAla..... 676
1460 ATGGAGCTGTCTCTCAGCGCTCCATCAGCTTGAGAGTAAGGATTAACC 1509
677 .....GlyLeuThrIleAspAlaLysAsnValThr 686
1510 CTCGAAACAGCCAGTGGATGATGGCACAGTGTGTGGACAGCACCGT 1559
687 ValAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 703
1560 GGGAAAGGACACTTGTCTTATCCTCGACACACGACGCTCCCCCAA 1609
703 rGlyGlu.....IleThrThrLysThr..... 710
1610 TCTTCTCTGGATCCAGTGGAGAGCAAGGTGGCTTTGTAGTCGAC 1659
711 .....GlyThrThrIleAsn 715
1660 AAAAACACCAAAATGGCTACCTCCAATCCAGCATTCCTAAGTTGG 1709
716 AlaThrThrGly.....AsnValGluIleThrAlaGlnThrGl 728
1710 CACTTGGAAATACAGTCTGCNAGCAAGCTCACAAACCTTGACCTGACTG 1759
729 ySerIleLeuGlyIleGluSerSerGlySerValThrLeuThrA 745
1760 TCAGTCCCGTGG.....TCCAATGTACCTCCCTGCCAATTTACA 1800
745 laThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 761
1801 GTGACTTCCAAAACAGCAAGACACACAGC..... 1830
762 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLy 778
1830 ..... 1830

```



```
778 sGlyThrGluSerValThrSerSerGlnSerGlnSerGlyAspIleGlyClyT 795
1831 .....AAATCCCCAGCCCTCTGGTAGTT 1854
795 hrIleSerGlyThrValGluValLysAlaThrGluSerLeuThrThr 811
1855 TATGCCAAATATTCGC.....CAAGGAGCTCCCAATCTCAG 1892
812 GlnSerAsnSerLysIleLysAlaThrGlyGluAlaAsnValThrSe 828
1893 GGCAGTGTACAGCCCTGATTGAATCACTGAATGAAAAACAGTTACCT 1942
828 rAlaThrGlyThr...IleGlyClyThrIleSerGlyAsnThrValAsnV 844
1943 TGGAA.....CTACTGGATAATGGACGAGGTGCTGAT 1974
844 aThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 860
1975 GCTACTAAG.....GATGACGGTGTCTACTCAAG 2003
861 AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 877
2004 GRATTTCAACACTTATGACACGAATGGTAGATACAGTGTAAAGTGGGG 2053
877 rGluAlaSerSerHisIleThrSerAlaLysGlyGlnValAsnLeuSerA 894
2054 CT.....CTGGAGGAGTTAAGCAGCCAGCGAGGAGTG 2088
894 laGlnAspGlySerValAlaGlySerIleAsnAlaAla.....AsnVal 908
2089 ATACCCACAGTGGAGCTGTACATACCTGGCTGGTATGAGAAATCA 2138
909 ThrLeuAsnThrThrGlyThrLeuThrThr.....ValLysGlySe 922
2139 TGAATATACAAATGGAATCCACCAAGACCTGAAATTAATAAGTATGTTT 2188
922 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 939
2189 AA.....CACAGCAAGTGTGTTTCAGCAGA 2214
939 luLeuAsnGlyAlaAlaLeuLeuGlyAsnHisThrValValAsnAlaThrAsn 955
2215 ACATCTCTGGGAGGCTATTGTTGGCTCTGATGTCCTCCAAATGCTCCCAT 2264
956 AlaAsnGlySerGlySerValIleAlaThrThrSerSerArgValAsnI 972
2265 ACCT...GATCTCTTCCACCTGGCCAAATC.....ACCG 2296
972 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleIleSerLysAsnG 989
2297 ACCTGAGCGGAAATTCAGGGGGCAGTCTCATTAATCTGACTTGGACA 2346
989 lyIleAsnThrValLeuLeuLysGlyValLysIleAspValLysIle 1005
2347 GCTCTCTGGG...GATGATTATGACCATGCAACAGCTCACAAATATATCAT 2393
1006 GlnProGlyIleAlaSerValAspGluValIleGluAlaLys..... 1019
2394 TCGAATAAGTACAGTATTCTTGTCTCAGACACAAAGTTCAATGATCTC 2443
1020 .ArgIleLeuGluLysValLysAspLeuSerAspGluGluArgGluAlaL 1036
2444 TT 2445
1036 eu 1036
```

seq_name: /SID52/gcgdata/geneseq/AA1993.DAT:AA41723

seq_documentation_block:

ID AA41723 standard; Protein: 1536 AA.

XX

AC

XX

XX

XX

XX

XX

XX

XX

XX

```
DT 20-APR-1994 (first entry)
XX
DE High molecular weight protein 1 (HMW1).
XX
FW HMW: high molecular weight protein; virus; vaccine; influenza;
KW epitope; immunity; haemophilus influenzae.
XX
OS Haemophilus influenzae.
XX
PN WO9319090-A.
XX
PD 30-SEP-1993.
XX
PF 16-MAR-1993; 93WO-US02166.
XX
PR 16-MAR-1992; 92GB-0005704.
XX
PA (BARE/) BARENKAMP S J.
XX
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Barenkamp SJ.
XX
XX
DR WPI: 1993-320683/40.
DR N-PSDB: AAQ49506.
XX
XX
PT High molecular weight surface proteins - of non-typeable
XX haemophilus which exhibit immunogenic properties
XX
PS Claim 3; Figure 2; 100pp; English.
XX
XX
CC The isolation and purification of the high molecular weight protein
CC enables the identification of the major protective epitopes of the
CC protein by conventional epitope mapping. These epitopes can then be
CC synthesised using standard techniques and incorporated into fully
CC synthetic or recombinant vaccines.
XX
XX
SQ Sequence 1536 AA;
```

alignment_scores:

```
Quality: 138.50 Length: 936
Ratio: 0.318 Gaps: 49
Percent Similarity: 46.581 Percent Identity: 20.406
```

alignment_block:

US-09-049-696-20 x AA41723

Align seg 1/1 to: AA41723 from: 1 to: 1536

```
217 TTGAAAGCTACA.....GGAAAGCGATTTTATTTCAAAATGT 254
|||||:|||||
631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnI 643
255 TGCCATTTTGATTCTCTGAAACATGGAAGACAAAGGCTGACTATGTGAGAC 304
|||||:|||||
643 eSerMetValLeuPro.....LysAsnGluSerGlyTyrAspLysP 657
305 CAAAAGTGTGAGACCTACAAAATGCTGATGTTCTG...CTTGTGAGTCT 351
|||||:|||||
657 heLysGlyArgThrTyrTrpAsnLeuThrSerLeuAsnValSerGluSer 673
352 .....ACTCTCCAGTAATGAT..... 369
|||||:|||||
674 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspSerAlaGlyTh 690
370 .....GAACCCCTACACT.....GAGCAGA 388
690 rLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThrT 707
389 TGGGCAACTGTGGAGAGA.....AGGGTGAAGGATCCACCTCCTCCT 432
|||||:|||||
707 hrPheAsnValGluArgAsnAlaArgValAsnPheAspIleLysAlaPr 723
```

1009 IleThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAspAlaThr 1025
::: |||:::||||:::||| ::|||::: |
1191 TCGATGGCCATTACTGTGATTAGGAAGA... 1221
:::||||| ::|||::: |||
1025 rAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnA 1042
:::||||| ::|||::: |||
1222TATCCAACCTGATGGA 1236
|||||
1042 sPLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysaspGly 1058
1237 TCTGAATTTGCTGCTGCAGGATGGGAACAACACACTATAAGTGG... 1284
||||| ::|||::: |||
1059 SerAsp.....LeuThrIleGlyAsnThrAsnSerAlaAspGlyTh 1072
:::||||| ::|||::: |||
1285TGCTTTAACGAGGTCAAACAAGT..... GTG 1312
|||||:::|||||:::|||||
1072 rAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089
1313 CCATCATCCACACAGCTGCTTTGGGCCCTCTGCAGCTCAAGAACCTAGAG 1362
||| ||| |||:::|||||
1089 laasPglyHisLysValThrLeuHis..... 1097
1363 GAGCTGTCCAAAATG...ACAGGAGGTTTACAGACATATGCTTCAGATCA 1409
|||||::: |||:::||||| ::|||::: |||
1098SerLysValGluThrSerGlySerAsnAsnThrLysAspSe 1112
1410 AGTTCAGAACATGGCCCTCATTTGATGCTTTTGGGGCCCTTTCATCAGGAA 1459
:::|||||::: |||
1112 rSerAspAsnAla..... 1117
1460 ATGGAGCTGTCCTCAGCGCTCCATCCAGCTTGAGACTAGGAGTTAAC 1509
::: |||::: |||::: |||::: |||
1118GlyLeuThrIleAspAlaLysAsnValThr 1127
1510 CTCCAGACAGCAGCTGGATGAATGGCACACTGATCTGGACAGCAGCGT 1559
:::||||::: |||::: |||
1128 ValAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrse 1144
1560 GGGAAGGACACTTTGTTCTATCATCTGCAGCAACAGCAGCTCCCCCAA 1609
|||||::: |||
1144 rGlyGlu.....IleThrLysThr..... 1151
1610 TCCTTCTCTGGATCCCAGTGCAGACAGCAAGCTGCTTAGTAGGGAC 1659
1152GlyThrThrIleAsn..... 1156
1660 AAAACACCACCAAAATGGCTTACCTCCAAATCCAGGCATTTGTAAGTTGG 1709
:::|||| ::|||::: |||::: |||
1157 AlaThrThrGly.....AsnValGluIleThrAlaGlnThrGl 1169
1710 CACTTGGAAATACAGTCTGCACAGCAAGCTCACAACTTGACCTGACTG 1759
ySerIleLeuGlyGlyIleGluSerSerGlySerValThrLeuThrA 1186
:::|||||::: |||
1760 TCAGTCCCGTGG.....TCCAATGTACCTGCCTCCCAATTACA 1800
:::||||::: |||
1186 lathrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
1801 GTGACTTCCAAAAACGAAGACACAGC..... 1830
|||||::: |||
1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLy 1219
1830..... 1830
1219 sglyThrGluSerValThrSerSerGlnSerGlyAspIleGlyGly 1236
1831AAATCCCCAGCCCTCTGTAAGT 1854
||| ::|||::: |||
1236 hrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThrThr 1252
1855 TATCCAATATTGCG.....CAAGGAGCTCTCCCAATTCAG 1892
:::||||::: |||

```
1253 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 1269
1893 GGCCAGTGTACAGCCCTGATTGAATCAGTGAATGGAACACAGTTACCT 1942
1269 rAlaThrGlyThr...ileGlyGlyThrIleSerGlyAsnThrValAsnV 1285
1943 TGGAA.....CTACTGGATAATGAGCAGAGTGTGAT 1974
1285 alThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 1301
1975 GCTACTAAG.....GATGACGGTGTCTACTCAAG 2003
1302 AlaThrGluGlyAlaAlaThrThrThrSerSerGlyLysLeuThrTh 1318
2004 GTATTTCACAACCTATGACAGCAATGGTACATACAGTGAAGTGGGG 2053
1318 rGluAlaSerSerHisIleThrSerAlaLysGlyGlnValAsnLeuSerA 1335
2054 CT.....CTGGAGGAGTTAAACGACGACGACGAGAGTG 2088
1335 laGlnAspGlySerValAlaGlySerIleAsnAlaAla.....AsnVal 1349
2089 ATACCCAGCAGAGTGGAGCAGTGTACATACCTGGCTGGTGAATGA 2138
1350 ThrLeuAsnThrThrGlyThrLeuThr.....ValLysGlySe 1363
2139 TGAATACAACTGAATGCCACAGAGCTGAAATTAATAAGATGATGTT 2188
1363 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 1380
2189 AA.....CACAGCAAGTGTGTTTCACGAGA 2214
1380 IuLeuAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn 1396
2215 ACATCTCGGGAGGCTCATTTGGGCTCTGATGCCAAATGCTCCCAT 2264
1397 AlaAsnGlySerGlySerValIleAlaThrThrSerSerArgValAsnII 1413
2265 ACCT...GATCTCTCCAGCTGCCAATC.....ACCG 2296
1413 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleIleSerLysAsnG 1430
2297 ACCTGAAGGGGAAATTCACGGGGGAGTCTCAATTAATCTGACTTGGACA 2346
1430 lylleAsnThrValLeuLeuLysGlyValLysIleAspValLysTyrIle 1446
2347 GCTCCTGGG...GATGATTATGACCATGGACAGCTCACAAAGTATATCAT 2393
1447 GlnProGlyIleAlaSerValAspGluValIleGluAlaLys..... 1460
2394 TCGAATAAGTACAACTATCTTGTATCTCAGAGACAAGTTCATGATCTC 2443
1461 .ArgIleLeuGluLysValLysAspLeuSerAspGluGluAlaL 1477
2444 TT 2445
1477 eu 1477
```

seq_name: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT:AA63505

seq_documentation_block:

ID AA63505 standard; Protein; 1536 AA.

XX AC AA63505;

XX DT 25-JUN-1995 (first entry)

XX DE Haemophilus high molecular weight protein HMW1.

XX KW High molecular weight protein; HMW1; protective vaccine; otitis;

XX KW sinusitis; bronchitis; Hib.

XX OS Haemophilus.

```
XX WO9421290-A.
XX 29-SEP-1994.
XX 15-MAR-1994; 94WO-US02550.
XX 16-MAR-1993; 93US-0038682.
XX (BARE/) BARENKAMP S J.
XX (SGEM/) ST GEME J W.
XX Ba'tenkamp SJ, St GEME JW;
XX WPI; 1994-316665/39.
XX Q-PSDB; Q72293.
XX New immunogenic high mol. wt. proteins of non typeable
XX Haemophilus - useful in protective vaccines
XX Claim 2; Page 31; 127pp; English.
XX The HMW1 protein encoded by this sequence is useful in a vaccine to
XX protect against disease caused by non-typeable Haemophilus which are
XX not controlled by H. influenzae type b (hib) vaccines. The encoded
XX protein can also be used as a carrier for protective Hib
XX polysaccharide (in a conjugate vaccin against meningitis) or for
XX other antigens, haptens, etc.
XX SQ Sequence 1536 AA;
```

alignment_scores:

Quality:	138.50	Length:	936
Ratio:	0.318	Gaps:	49
Percent Similarity:	46.581	Percent Identity:	20.406

alignment_block:

US-09-049-696-20 x AAR63505

Align seg 1/1 to: AAR63505 from: 1 to: 1536

```
217 TTGTGAAGCTACA.....GGAAACCGATTTTATTTCAAAAATGT 254
|||||:|||||
631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnII 643
255 TGCATTTTGTATCTCTGAACATGGAAGACAAAGCTGACTATGTGAGAC 304
|||||:|||||
643 eSerMetValLeuPro.....LysAsnGluSerGlyTyrAspLysp 657
305 CAAAACCTGAGACCTCAAAAATGCTGATTTCTG...GTTGCTGAGTCT 351
|||||:|||||
657 helysGlyArgThrTyrTrpAsnLeuThrSerLeuAsnValSerGluSer 673
352 .....ACTCCTCCAGGTAATCAT..... 369
|||||:|||||
674 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspSerAlaGlyTh 690
370 .....GAACCCCTACACT.....GAGCAGA 388
|||||:|||||
690 rLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThr 707
389 TGGGCAACTGTGGAGAGA.....AGGGTGAAAGGATCCACCTCACTCT 432
|||||:|||||
707 hrPheAsn.ValGluArgAsnAlaArgValAsnPheAspIleLysAlaPr 723
433 GATTTCATTCACGAAAAAGTTAGCTGAATATGGACCACCAAGTAGGGC 482
|||||:|||||
723 oIleGlyIleAsnLysTyrSerSerLeuAsnTyrAlaSerPheAsnGlyA 740
483 ATTTGTCATGAGTGGGCTCATCTACGATGGGAGTATTTCACGAGTACA 532
|||||:|||||
740 snIleSerValSerGly.....GlyGlySerValAspPheThr 752
```



```

1285 aThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 1301
1290 .....
1295 GCTACTAAG.....GATGACGGTCTCTACTCAAG 2003
1300 AlAThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 1318
1305 .....
1310 GTATTTCACAACTATGACACGAGTGTAGATACAGTGTAAAGTGGGG 2053
1315 rGluAlaSerSerHisIleThrSerAlaLysGlyGlnValAsnLeuSerA 1335
1320 .....
1325 CT.....CTGGGAGGAGTAAACGACGACGACGAGGAGTG 2088
1330 .....
1335 laglnAspGlySerValAlaGlySerIleAsnAlaAla.....AsnVal 1349
1340 .....
1345 ATACCCACGACGAGTGGACGACTGTACATCCGTGGCTGATGAGAAATGA 2138
1350 ThrLeuAsnThrThrGlyThrLeuThrThr.....ValLysGlySe 1363
1355 .....
1360 TGAATACATGAATCCACCAAGCTGNAATTAATAGGATGATGTTTC 2188
1365 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 1380
1370 .....
1375 AA.....CACAAGCAAGTGTGTTTCACGAGA 2214
1380 lueAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn 1396
1385 .....
1390 ACATCTCGGAGGCTCATTTGGCTCTGTATGTCCTCCAAATGCTCCCAT 2264
1395 .....
1400 AlaAsnGlySerGlySerValIleAlaThrThrSerSerArgValAsnIl 1413
1405 .....
1410 ACCT...GATCTCTCCACTGGCCCAATC.....ACCG 2296
1415 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleIleSerLysAsnG 1430
1420 .....
1425 ACCTGAAGCGGAAATTCACGGGGGAGCTCTAATTAATCTGCTTGGACA 2346
1430 lyleAsnThrValLeuLeuLysGlyValLysIleAspValIleAspIle 1446
1435 .....
1440 GTCCTCTGGG...GATGATTATGACCATGGAACGCTCACAAATGATATCAT 2393
1445 .....
1450 GlnProGlyIleAlaSerValAspGluValIleGluAlaLys..... 1460
1455 .....
1460 TCAATTAAGTACAAGTATCTGATCTCAGACACAGTTCATGATCTC 2443
1465 .....
1470 ArgIleLeuGluLysValLysAspLeuSerAspGluGluArgGluAlaL 1477
1475 .....
1480 TT 2445
1485 eu 1477

```

seq_name: /SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAB01846

seq_documentation_block:

ID AAB01846 standard; Protein; 1536 AA.

XX AAB01846;

AC AAB01846;

DT 11-SEP-2000 (first entry)

DE Haemophilus influenzae strain 12 HMW1A protein, SEQ ID NO:67.

KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW recombinant production; Escherichia coli; non-encapsulated;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.

OS Haemophilus influenzae strain 12.

PN WO200020609-A2.

PD 13-APR-2000.

```

XX 07-OCT-1999; 99WO-CA00938.
XX 07-OCT-1998; 98US-0167568.
XX 08-DEC-1998; 98US-0206942.
XX (CONN-) CONNAUGHT LAB LTD.
XX Loosmore SM, Yang Y, Klein MH;
XX WPI; 2000-303789/26.
XX N-PSDB; AAA52195.
XX Nucleic acid molecule for producing recombinant high molecular weight
XX proteins of Haemophilus which are used as a vaccine to provide
XX protection against Haemophilus induced diseases in humans -
XX Example 16; Fig 28A-Q; 307pp; English.
XX The invention relates to the recombinant production of Haemophilus
XX influenzae high molecular weight (HMW) proteins in Escherichia coli. The
XX expression construct used to effect recombinant expression comprises a
XX promoter functional in E. coli (e.g., the T7 promoter) operably linked
XX to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
XX influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene
XX clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
XX hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
XX and the hmwB and hmwC genes encode accessory proteins which are
XX responsible for post-translational processing and secretion of the HMW
XX proteins. The modified hmwABC operon used in the expression construct of
XX the invention contains an A gene modified such that it encodes only the
XX mature HMW. The invention also discloses hmwA genes (AAA52175-A52198)
XX and HMW proteins (AAB01824-B01849) from the non-typeable H. influenzae
XX strains Joyn, K1, K21, LDC2, PMH1, 15 and 12. The nucleic acids and
XX vectors are used for the production of recombinant H. influenzae HMW
XX proteins which can be used as vaccines to mediate a humoral or
XX cell-mediated immune response to provide protection against diseases in
XX humans caused by H. influenzae (e.g., otitis media, epiglottitis,
XX pneumonia and tracheobronchitis). The HMW proteins are also useful as
XX antigens in immunoassays for detecting antibodies against Haemophilus,
XX HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
XX HMW proteins can be used to isolate and clone hmw genes from other
XX non-typeable strains of Haemophilus via hybridisation reactions. The
XX present sequence represents an HMW protein from a non-typeable strain of
XX H. influenzae.
XX SQ Sequence 1536 AA;

```

alignme:t_scores:
Quality: 138.50 Length: 936
Ratio: 0.318 Gaps: 49
Percent Similarity: 46.581 Percent Identity: 20.406

alignment_block:
US-09-049-696-20 x AAB01846 ..

Align seg 1/1 to: AAB01846 from: 1 to: 1536

```

217 TTTGAGCTACA.....GGAAGCGATTTTATTTCAAAATGT 254
|||||:|||||
631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnIl 643
255 TGCCATTTTGATTCTCCGAAACATGGAAGACAAAGCGTGTGCTGAGAC 304
|||||:|||||
643 eSerMetValLeuPro.....LysAsnGluSerGlyTyAspLysP 657
305 CAAACTTGAGACCTACAAAATGCTGATGTTCTG...GTTGCTGAGTCT 351
|||||:|||||
657 heLysGlyArgThrTyTrpAsnLeuThrSerLeuAsnValSerGluSer 673
352 .....ACTCCTCCAGGTAATGAT..... 369
|||||:|||||

```

674 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspSerAlaGlyTh 690
370GAACCTACACTGAGCAGA 388
690 rLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThr 707
389 TGGCAACTGTGGAGAGA.....AGGCTGAAAGGATCCACCTCACTCCT 432
707 hrPheAsn.ValGluArgAsnAlaArgValAsnPheAspIleLysAlaPr 723
433 GATTTCATTCGAGGAAAGAGTTAGCTGATATAGGACCACAAAGTAGGCG 482
723 oileGlyIleAsnLysTyrSerLeuAsnTyrAlaSerPheAsnGlyA 740
483 ATTTCATTCAGTGGGCTCATCTACGATGGGAGTATTGACGAGTACA 532
740 snIleSerValSerGlyGlyGlySerValAspPheThr 752
533 ATAATGATGAGAAATTCATTATCCAAATGGAAGATACA.....AGCA 576
753 LeuLeuAlaSerSerSerAsn.ValGlnThrProGlyValIleAsnS 769
577 GTAGATGTTGACGAGTATTACTGGTACAAATGTAGTAAAGAGTGTCA 626
769 erLysTyrPheAsn.valSerThrGlySerSerLeuArgPheLysThrSe 785
627 GGGG.....GGCAGCTGTTACACCACAAAGATGCACATTCAT. 663
785 rGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeuAsnA 802
664AACTACAGCAGCTCTATGAAAGAGGATGTAGTTGTT 702
802 laThrGlyAsnIleThrLeuLeuGlnValGluGlyThrAspLysMet 818
703 CTCCAATCCGCGACGAGGAGGCTTCTATAATGTT 741
819 ileGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyGlyAs 835
742GCACACATGTTGATTCTATAGTTG 766
835 nileThrPheGlySerArgLysAlaValThrGluIleGluGlyAsnVal 852
767 AATTCGTACAGAACAAAAC 786
852 hrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPheAspAsn 868
787 CACAACAAGAGCTCCAAACAGCAAAATCAAAATGCCAA 827
869 HisGlnLysProLeuThrIleLysLysAspValIleIleAsnSerGlyAs 885
828 TCTCCGACCATGGGAGTGATCC 853
885 nLeuThrAlaGlyLysIleValAsnIleAlaGlyAsnLeuThrValG 902
854 GTGATTCTCAGGACTTTAAGAAACACCTCCATGACA 891
902 luserAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsnValGly 918
892ACACAGCCCAATCCACCTTCTCATTCGTCGAGATTGG 932
919 GlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLysGlyGl 935
933 ACNAAAGATTGTGTTAGTCTTCACAAATCTGGAAGCATGGGACTG 982
935 yAlaArgPheLysAspIleAspAsnSerLysAsnLeuSerIleThrThra 952
983 GTAACCGCCTCAATCGACTGAATCAA 1008
952 snSerSerThrTyrArgThrIleIleSerGlyAsnIleThrAsnLys 968
1009 GCAGGCCAGCTTTCTGCTGCAGACAGTTGAGCTGGGTCTCTGGTTGG 1058
969 AsnGlyAspLeuAsnIleThrAsnGluGlySer 979

1059 GATGTGACATTGTGACAGTGTGCCCATCTACAAAGTGAACCTCATACAG. 1107
980AspThrGluMetGlnIleGlyGlyAspValSerGlnL 992
1108ATAACAGTGCAGTGCAGGACACACACTCCCAAAAGA 1146
992 ysGluGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrLysGln 1008
1147TTACCTGCAGCAGCTTCAGAGGAGCTCATCTGCAGCGGCT 1190
1009 IleThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAlaTh 1025
1191 TCGATCGCATTTACTGCTATTAGGAAGAAA 1221
1025 rAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnA 1042
1222TATCCAACATGATGGA 1236
1042 sPleuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGly 1058
1237 TCTGAAATTGCTGCTGACGAGTGGGAAGACACACTATAGTGGG... 1284
1059 SerAsp.....LeuThrIleGlyAsnThrAsnSerAlaAspGlyTh 1072
1285TGCTTTAACGAGGTCAACAAAGT.....GGTG 1312
1072 rAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089
1313 CCATCATCCACACAGTGCCTTTGGGCCCTCTGCAGCTCAAGAACTAGAG 1362
1089 laAspGlyHisLysValThrLeuHis 1097
1363 GACGTGTCAAAATG...ACAGGAGTGTTCACACATATGCTTCAGATCA 1409
1099SerLysValGluThrSerGlySerAsnAsnAsnThrGluAspSe 1112
1410 AGTTCAGAACAAATGGCCTCATGATGCTTTGGGGCCCTTTTCATCAGGA 1459
1112 rSerAspAsnAsnAla 1117
1460 ATGAGCTGCTCTCAGCGCTCCATCCAGCTTCAGAGTGAAGGATTAACC 1509
1118GlyLeuThrIleAspAlaLysAsnValThr 1127
1519 CTCGAGAACAGCCAGTGGATGAATGGCAGATGATCGTGACAGCACCGT 1559
1128 ValAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144
1560 GGGAAAGGACACTTTGTTCTTATCCTCGACACAGCAGCCTCCCAAAA 1609
1144 rGlyGluIleThrThrLysThr 1151
1610 TCCTTCTCTGGGATCCAGTGGAGAGCAAGGAGTGGCTTTGTAGTGGAC 1659
1152GlyThrThrIleAsn 1156
1660 AAAACACAAAATGGCTACCTCCAATCCAGGCATTGCTAAGGTGG 1709
1157 AlaThrThrGlyAsnValIleThrAlaGlnThrGl 1169
1710 CACTTGGAAATFACAGTCTGCAAGCAAGCTCACAAACCTTGACCTGACTG 1759
1169 ySerIleLeuGlyGlyIleGluSerSerGlySerValThrLeuThrA 1186
1760 TCAGTCCCGCTGGCTCCAATGCTACCTGCCTCCCAATTACA 1800
1186 laThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
1801 GTGACTTCCAAAACGACAAAGACAGCACCCAGC 1830
1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLy 1219

```
1830 ..... 1830
1219 sglyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyGlyT 1236
1831 ..... AAATTCGCCAGCCCTCTGGTAGTT 1854
1236 hrIleSerGlyThrValGluValIleAlaThrGluSerLeuThrThr 1252
1855 TAGCAATATTTCGC ..... CAAGGAGCCTCCCAATTCCTAG 1892
1253 GlnSerAsnSerIleIleAlaThrThrGlyGluAlaAsnValThrSe 1269
1893 GGCACGTCTGCAGCCCTGATTCAATCACTGAATGCAAAACAGTTTACCT 1942
1269 rAlaThrGlyThr...IleGlyGlyThrIleSerGlyAsnThrValAsnV 1285
1943 TGGAA .....CTACTGGATAATGAGCAGGTGCTGAT 1974
1285 alThrAlaAsnAlaGlyAspLeuThrValIleGlyAsnGlyAlaGluIleAsn 1301
1975 GCTACTAAG .....GATGACGGTGTCTACTCAAG 2003
1302 AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 1318
2004 GTATTTCAACAATTATCACACGATGCTAGATACACTGTAAGAGTCGGG 2053
1318 rGluAlaSerSerHisIleThrSerAlaLysGlyGlnValAsnLeuSerA 1335
2054 CT .....CTGGAGGAGGTAAACGCAGCAGCAGGAGGTG 2088
1335 laGlnAspGlySerValAlaGlySerIleAsnAlaAla ..... AsnVal 1349
2089 ATACCCACGAGAGTGAGCTGACACTGACCTGCTGGATTGAGAAATGA 2138
1350 ThrLeuAsnThrThrGlyThrLeuThrThr .....ValLysGlySe 1363
2139 TGAATACAAATGAATCCACCAAGACCTGAATTAATAAGGATGATGTC 2188
1363 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 1380
2189 AA .....CACAGCAAGTGTGTTTCAGCAGA 2214
1380 luLeuAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn 1396
2215 ACATCTCGGAGGCTCATTTGGCTTCTGATGTCCTCCAAATGCTCCCAT 2264
1397 AlaAsnGlySerGlySerValIleAlaThrThrSerSerArgValAsnI 1413
2265 ACCT...GATCTCTCCACCTGGCCAAATC .....ACCG 2296
1413 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleIleSerLysAsnG 1430
2297 ACCTGAAGCGGAATATTCAGGGGCGAGTCTCATTAATCTGACTTGGACA 2346
1430 lyIleAsnThrValLeuLeuLysGlyValLysIleAspValLysTyrIle 1446
2347 GCTCCTGGG...GATGATTATGACCATGGAACAGCTCACAAAGTATATCAT 2393
1447 GlnProGlyIleAlaSerValAspGluValIleGluAlaLys ..... 1460
2394 TCGAATAAGTACAGATATTCTTCATCTCAGACAGCAAGTTCAATCAATCTC 2443
1461 .ArgIleLeuGluLysValLysAspLeuSerAspGluGluArgGluAlaL 1477
2444 TT 2445
1477 eu 1477
```

seq_name: /SID2/gcgdata/geneseq/geneseq/AA2001.DAT: AAG89860

seq_documentation_block:

ID AAG89860 standard; Protein; 1152 AA.

XX

```
AC AAG89860;
XX
XX 26-SEP-2001 (first entry)
XX C glutamicum-protein fragment SEQ ID NO: 3614.
XX
XX CoIyneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX CoIynebacterium glutamicum.
XX EPI1108790-A2;
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
XX 07-APR-2000; 2000JP-0159162.
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
XX N-PSDB; AAH65079.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX
XX Claim 17; SEQ ID NO: 3614; 245pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a protein described
XX in the exemplification of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX European Patent Office.
XX
XX Sequence 1152 AA;
XX
XX alignment_scores:
XX Quality: 135.50 Length: 801
XX Ratio: 0.383 Gaps: 38
XX Percent Similarity: 44.195 Percent Identity: 19.476
XX
XX alignment_block:
XX US-09-049-696-20 x AAG89860 ..
XX
XX Align, seg 1/1 to: AAG89860 from: 1 to: 1152
XX
XX 55 TTGATTCTTCACCTCTAGAAGGGCGCCTGAGTAATTCACCTATTGAGCT 104
XX ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 478 LeuValProPheIleAsnGluProAspThrAsnProAlaLeuIleGlnAs 494
XX
XX 105 GAACAACAATGGCTATGAGGCAATTCGTTGCAATCGACCCCAATGTC 154
XX : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 494 nGlnGlnTrpAlaThrLysAlaValAlaThrAlaGluProGlyTrpL 511
XX
XX 155 CAGAAGATGAACA ..... 168
XX ||| : : : |||
```

511 euGluLysGlnThrLysProGluValLeuGluGluGlyAspValAspLys 527
169 CTCATTCAACAAATAAGGACATGGTACACCGAGCATCTCTGTATCTGTT 218
528 LeuIleAsnAspValArgAspAlaAla..... 536
219 TGAAGCTACAGAAAGCGATTTATTTCAAAATGTGGCCATTGTGATTC 268
536 536
269 CTGAACATGGAACACAAAGGCTGACTATGTGAGACCAAACTGTGAGACC 318
537 ..GluAlaThrAlaAlaArgProAlaArgGluArgAlaGluIle...Leu 551
319 TACAAAATGCTGATGTTCTGTTGCTGAGTCTACTCTCCAGTAATGA 368
552 TyrLysThrAlaGluIleLeuArgVal..... 560
369 TGAACCTTACACTGACGAGATGGCAACTGTGGAGAGAAGGGTGAAGGA 418
561ArgArgGlyHisLeuI 566
419 TCCACCTCACTCTGATTTCATTGCAGCAAAAAGTTAGCTGAATATGGA 468
566 leSerValThrAlaAlaGluVal...GlyLysAlaValGluGlnThrAsp 581
469 CCA.....CAAGGTAGGCGCATTTGTCATGAGTGGGCTCATCTACG 509
582 ProGluIleSerGluAlaIleAspPheAlaArgTyrTyrAlaHisLeuAl 598
510 ATGGGGAGTATTGAC.....GAGTACAAATATGATGAGAAT 547
598 aLeuGluLeuAspAspValAspAlaGluPheThrProAspArgValV 615
548 TCTACTTATCCAAATGAAGATACAGCAGTAAGATGTCACGAGGTATT 597
615 alValValThrProThrPsnPheProIleAlaIleProAlaGlySer 631
598 ACT.....GGTCAAAATGTAGTAAAGAAGTGTACGGG 629
632 ThrPheAlaAlaLeuAlaGlyAlaGlyValIleHisLys..... 645
630 AGGAGCTGTACACCAAGATGACATTCNATAAGTAACAGCAGCTCT 679
646 .ProSerLysProSerGlnHisCysSerAlaAlaValValGluAlaLeuT 662
680 ATGAAAAAGGA.....TGTGAGTTTGTCTCCAAATCCCGC 714
662 rpGluAlaGlyValProArgGluValLeuHisCysIleTyrProAlaAsn 678
715 CAGACGGAGAGGCTTCTATAATGTTTGCACAAATGTTGATTCATAGT 764
679 ArgAspValGlyCysAlaLeuIleSerHisGluHisValAspArgValI 695
765 TGAATTCTGTACAGCAAAACCAACCAAGAGCTCCAAACAGCAAA 814
695 e.....LeuThrGlySerS 700
815 ATCAAAATGCAATCTCCGACACACATGGGAAGTATCGGTGATTCGTAG 864
700 erGluThrAlaAlaMetPheSerSerTrpArg..... 710
865 GACTTTAAGAAACCACTCTATGACACACAGCCACCAAAATCCACCTT 914
711ProGluLeuThrI 715
915 CTCACTGTCGAGATTGGACAAGAATTTGTGTTTAGTCTTGACAAAT 964
715 easnGlyGluThrSerGlyLysAsnAlaIleValValThrProSerAlaA 732
965 CTGGAAGCATGGGACTGTAAACCGCTC...AATCGACTGAATCAAGCA 1011
732 spArgAspLeuAlaValAlaAspLeuValLysSerAlaPheGlyHisAla 748

1612 GCCAGCTTTTCTCTGTCGACAGACTTGAGCTGGGGTCTCGGTTGGGAT 1061
745 GlyGln.....LysCysSerAlaAlaSerLeuGlyIleLeuValGlySe 763
1062 GTGACATTTGACAGTGTGCCCACTGACAAAAGTGAACATCATACAGATAA 1111
763 rValTyr.....GluSerGluArgPheArgLysGlnLeuValAspAlaA 778
1112 ACAGTGGCAGTGACAGGGACACACTGCCAAAAGATTACTGCAGCAGCT 1161
778 laSerSerLeuIleValAspTrpProThrAsn.....ProSerAlaThr 792
1162 TCAGGAGGG...ACGTCCATCTGCAGCGGG...CTTCGATCGCATTTAC 1205
793 ValClyProLeuThrGluLeuProSerAspLysLeuHisHisAlaLeuTh 809
1206 TGTGATTAGGAAGAAATATCCAACTGATGATCTGAAATTTGCTGCTGA 1255
809 rThrLeuGlu.....GluGlyGluSerTrpLeuLeuLysP 821
1256 CGGATGGGGAAGACAACT.....ATAAGTGGTGGTCTTTAACGAG 1296
821 roArgGlnLeuAspAspThrGlyArgLeuTrpSerProGlyIleLysGlu 837
1297 GTCAACAAAGTGTGCCATCATCCACACAGTC.....GC 1331
838 GlyValLysProGlyThrPhePheHisLeuThrGluValPheGlyProva 854
1332 TTTGGGCGCTCTGACAGTCAAGAAGTAGAGAGCTGTCCAAA..... 1374
854 lLeuGlyLeuMetLysAlaThrAspLeuAsnGluAlaIleGluPheGlnA 871
1375ATCACAGGAGGTTTACAGACATATGCTTCAGAT 1407
871 snGlyAsnAspPheGlyLeuThrGlyGlyLeuGlnSerLeuAspAlaAsp 887
1408 CAAGTTTCAG.....AA 1418
888 GluValArgThrTrpLeuAspHisValAspValGlyAsnAlaTyrValas 904
1419 CAATGGCTCATTTGATCTTTGGGCGCTTTCATCAGGAAATGGAGCTG 1468
904 naArgGlyLe.....ThrGlyAlaI 911
1469 TCTCTAGCGCTCCATCCAGCTGACAGTAAAGGATTAACCCCTCCAGAAC 1518
911 leValGlnArg.....GlnSerPheGly..... 918
1519 AGCCAGTGGATGAATGGACAGTATCGTGACAGCACCGCTGGGAAAGGA 1568
919 ...GlyTrpLysLysSerSerValGlyLeuGlySerLysAlaGlyGlyPr 934
1569 CACTTTGTTCTTATC.....ACCTGGACACCCGACCT..... 1602
934 oAsnTyrValMetLeuMetGlyThrTrpAlaAspAlaProSerHisAla 951
1603 ..CCCCAAATCCTTCTCTGGGATCCAGTGGACAGCAAGAGGTGGCTTT 1650
951 laProArg..... 953
1651 GTAGTGGACAAAAACACCAAAATGGCTACTCCAAATCCAGGCATTCG 1700
954GluThrAsnProLeuIleSerLysLeuAspLeuProGly..... 966
1701 TAAGTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAAACCTTGA 1750
967 .GluGluLeuGluThrLeuGluLysAlaAsnAlaSerAspGluThr.... 981
1751 CCCTGACTGTCACTGCCGTCCGCTCAATGCTACCCCTGCCCTCAATTACA 1800
981

SQ Sequence 1152 AA; alignment_scores: Quality: 135.50 Length: 801
Ratio: 0.383 Gaps: 38
Percent Similarity: 44.195 Percent Identity: 19.476
alignment_block:
US-09-049-696-20 x AAB79787 ..
Align seg 1/1 to: AAB79787 from: 1 to: 1152
55 TTGATTCTTCACTTCTAGAGGGCCCTGAGTAATTCACCTACCTCAGCT 104
478 LeuValProPheIleAsnGluProAspThrAsnProAlaLeuIleGlnAs 494
105 GAACAACAATGGCTATGAAGGCATTGCTTGCATTCGACCCCAATATGC 154
494 nGlnGlnTrpAlaThrLysAlaValAlaThrAlaAlaGluProGlyTrpL 511
155 CAGAAGATGAACA..... 168
511 euGluLysGlnThrLysProGluValLeuGluGluGlyAspValAspLys 527
169 CTCATTCAACAATAAAGGACATGGTGACCCAGGCATCTCTGTATCTGTT 218
528 LeuIleAsnAspValArgAspAlaAla..... 536
219 TGAAGCTACAGGAAGCGATTTTATTTCAAAAATGTTGCCATTTGATTC 268
536 536
269 CTGAACATGGAAGACAAGGCTGACTATGTGAGACAAACTTGAGACC 318
537 ..GluAlaTrpAlaAlaArgProAlaArgGluArgAlaGluIle...Leu 551
319 TACAAAAATGCTGATGTTCTGCTGCTGAGTCTACTCTCCAGGTAATGA 368
552 TyrLysThrAlaGluIleLeuArgVal..... 560
369 TGAACCCCTACACTGACGAGATGGGCACTGTGGAGAGAGGTGAAGGA 418
561ArgArgGlyHisLeuI 566
419 TCCACCTCACTCCTCATTTTCATTCGAGGAAAAAAGTTAGCTCAATATGA 468
566 leSerValThrAlaAlaGluVal...GlyLysAlaValGluGlnThrAsp 581
469 CCA.....CAAGGTAGGCATTTGTCCTGAGTGGGCTCATCTACG 509
582 ProGluIleSerGluAlaIleAspPheAlaArgTyrTyrAlaHisLeuAl 598
510 ATGGGGAGTATTGAC.....GAGTACAATAATGATGAGAAAT 547
598 aLeuGluLeuAspValAspAlaGluPheThrProAspArgValV 615
548 TCTACTTATCCCAATGAGCAATACAGCAGTAAGATGTTTCACAGGATTT 597
615 aValValThrProProTyrAsnPheProIleAlaIleProAlaGlySer 631
598 ACT.....GGTACAATGTAGTAAGAAGTGTGAGG 629
632 ThrPheAlaAlaLeuAlaAlaGlyAlaGlyValIleHisLys..... 645
630 AGGCAGCTGTACACCAAAAGTGCACATTCATTAAGTAACAGGACTCT 679
646 .ProSerLysProSerGlnHisCysSerAlaAlaValGluAlaLeuT 662
680 ATGAAAAAGA.....TGTGAGTTTGTCTTCTCCAAATCCCGC 714
662 rpgluAlaGlyValProArgGluValLeuHisCysIleTyrProAlaAsn 678

715 CAGACGGAGAGGCTTCTATATATGTTTGCACAAACATGTTGATTTATAGT 764
675 ArgAspValGlyCysAlaLeuIleSerHisGluHisValAspArgValI 695
765 TGAATTCGTACAGAACAAACCAACAAAGAAAGCTCCAAACAGCAAA 814
695.e.....LeuThrGlySerS 700
815 ATCAAAAATGCAATCTCCGAAGCAGCATGGGAAGTGCATCCGTGATTCTGAG 864
700 erGluThrAlaAlaMetPheSerSerTrpArg..... 710
865 GACTTTAAGAAAACCTCTCTATGACACACAGCCACCAAAATCCACACTT 914
711ProGluLeuThrI 715
915 CTCATTGCTGCAGATTGACAAAAGAAATTCGTGTTTAGTCCTTGACAAAT 964
715 eaSnGlyGluThrSerGlyLysAsnAlaIleValValThrProSerAlaA 732
965 CTGGAAGCATGGCGACTGGTAACCGCTC...AATCGACTGAATCAAGCA 1011
732 spArgAspLeuAlaValAlaAspLeuValLysSerAlaPheGlyHisAla 748
1012 GGCACGCTTTCTCTGCTGCAGACAGTTGAGCTGGGTCTGCTGGTGGAT 1061
749 GlyGln.....LysCysSerAlaAlaSerLeuGlyIleLeuValGlySe 763
1062 GGTGACATTTGACAGTCTGCCCATGTACAAAAGTGAACATCATAAGATAA 1111
763 rValTyr.....GluSerGluArgPheArgLysGlnLeuValAspAlaA 778
1112 ACAGTGGCAGTGACAGGACACACTGCCAAAAGATTACTCCAGCAGCT 1161
778 laSerSerLeuIleValAspTrpProThrAsn.....ProSerAlaThr 792
1162 TCAGGAGG...ACGTCCATCTGCAGGG...CTTCGATCGCATTTAC 1205
793 valGlyProLeuThrGluLeuProSerAspLysLeuHisAlaLeuTh 809
1206 TGTGATTAGGAAGAAATATCCAACTGATGATCTGAATGTTGCTCTCTGA 1255
809 rThrLeuGlu.....GluGlyGluSerTrpLeuLysP 821
1256 CGGATGGGAGACAACT.....ATAAGTGGTCTTTTAACGAG 1296
821.roArgGlnLeuAspAspThrGlyArgLeuTrpSerProGlyIleLysGlu 837
1297 GTCAAACAAAGTGTGCCATCATCCACACAGTC.....GC 1331
838 GlyValLysProGlyThrPhePheHisLeuThrGluValPheGlyProVa 854
1332 TTGCGGCGCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAA..... 1374
854 lLeuGlyLeuMetLysAlaThrAspLeuAsnGluAlaIleGluPheGlnA 871
1375ATCAGAGGAGGTTTACAGACATATGCTTTCAGAT 1407
871 snGlyAsnAspPheGlyLeuThrGlyGlyLeuGlnSerLeuAspAlaAsp 887
1408 CAAGTTTCAG.....AA 1418
888 GluValArgThrTrpLeuAspHisValAspValGlyAsnAlaTyrValAs 904
1419 CAATGGCGCTCATTCATGCTTTTGGGCGCTTTCATCAGGAAATGGAGCTG 1468
904 nArgGlyIle.....ThrGlyAlaI 911
1469.TCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAACCTCCAGAAC 1518
911 leValGlnArg.....GlnSerPheGly..... 918
1519 AGCCAGTGGATGAATGGCACAGCTGATCGTGACAGACCCCGTGGGAAGGA 1568

```
919 ...GlyTrpLysSerSerValGlyLeuGlySerLysValGlyGlyPr 934
1569 CACTTTGTTCTTATC.....ACCTGGACAACGACGCTT.....1602
...:.....:.....:.....:.....:.....:.....:.....:.....:
934 oAsnTyrValMetLeuMetGlyThrTrpAlaAspAlaProSerHisHis 951
...:.....:.....:.....:.....:.....:.....:.....:.....:
1603 ..CCCCAATCCTCTCTGGGATCCAGTGGAGGAGAGCANGGTGGCTTT 1650
...:.....:.....:.....:.....:.....:.....:.....:.....:
951 laProArg.....:.....:.....:.....:.....:.....:.....: 953
1651 GTAGTGGACAAAACACCAAAATGGCTACTCCAAATCCAGGCAATGTC 1700
...:.....:.....:.....:.....:.....:.....:.....:.....:
954 .....GluThrAsnProLeuIleSerLysLeuAspLeuProGly..... 966
1701 TAAGTTGGCACTTGGAAATACACTGTGCAAGCAAGCTCACAAACCTTGA 1750
...:.....:.....:.....:.....:.....:.....:.....:.....:
967 .GluGluLeuGluTrpLeuGluLysAlaAsnAlaSerAspGluThr.... 981
1751 CCCTGACTGTCAGTCCCGTGGCTCCCAATGCTACCCCTGCCCTCCCAATTACA 1800
981 .....:.....:.....:.....:.....:.....:.....:.....: 981
1801 GTGACTTCCAAAACGACAAAGGACACCAAAATCCAGCCCTCTGGT 1850
...:.....:.....:.....:.....:.....:.....:.....:.....:
982 ...AlaTrpAsnThrGluPheGlySerProArgAspProSerGlyLeuAs 997
1851 AGTTTATGCAATATT.....:.....:.....:.....:.....:.....: 1866
...:.....:.....:.....:.....:.....:.....:.....:.....:
997 pValGluAlaAsnIlePheArgTyrArgProAlaGluValValLeuArgL 1014
1867 .....:.....:.....:.....:.....:.....:.....:.....: 1890
1014 euAspAspSerAlaThrProArgGluThrAlaArgAlaLeuLeuAlaA 1030
1891 ...AGGCCAGTGTACAGCCCTGATTGAATCAGTGAATGGAAAAACAGT 1937
1031 ArgArgAlaGlyValThrProArgValLeuGlnThrProGlyValSerG 1047
1938 TACCTTGGAACTACTGATGAAATGAGCAGTGTGTGATGCT...ACTAAGG 1984
1047 uGlnValArgGluValLeuSerAlaAlaGlyValSerAlaGluThrVal 1064
1985 ATGACGGTGTCTACTCAAGTATTTCACA.....ACTTATGACACGAAT 2028
1064 spAspSerValPheIleSerAsnValLeuArgGlyGluTyrAspGluAsn 1080
2029 GGTAGATACAGTGTAAAGTGCAGGCTCTGGAGGAGTTAAAC..... 2070
1081 Ser.....SerValArgValArgTyrLeuGlyLysValSerAspThrVa 1095
2071 .GCAGCCAGACGGAGAGTGTATACCCAG.....:.....:.....: 2097
1095 la:rgGluArgLeuSerValArgProGluValValLeuLeuAspAlav 1112
2098 .....CAGACTGGGCACTGTACATCCTGGCTGGATTGAGATGATGAA 2142
1112 alThrAlaSerGlyArgValGluLeuArgTyrTrpLeuLysGluGlnAla 1128
2143 ATA 2145
1129 Ile 1129
```

seq_name: /SID2/gcgdata/geneseq/geneseq/AA2001.DAT:AAU05396

seq_documentation_block:

ID AAU05396 standard; Protein; 26926 AA.

XX

AC AAU05396;

XX

DT 24-OCT-2001 (first entry)

XX

DE Human titin (connectin) protein sequence.

```
XX Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;
FW titin-related disease; zebrafish; heart failure; heart disease.
XX Homo sapiens.
XX WO200151666-A1.
XX 19-JUL-2001.
XX 12-JAN-2001; 2001WO-US01212.
XX 12-JAN-2000; 2000US-0175787.
XX (GEO ) GEN HOSPITAL CORP.
XX Fishman MC;
XX WPI; 2001-451869/48.
XX N-P5DB; AAS05390.
XX
XX Determining if a subject has or is at risk of developing a
XX titin-related disease or condition, particularly heart failures,
XX comprises detecting the presence of a mutation in the titin gene
XX
XX Disclosure; Page 57-111; 114pp; English.
```

```
XX The present sequence representing human titin (also known as connectin)
XX is described in an invention relating to a novel method for determining
XX whether a subject has or is at risk of developing a titin-related
XX disease or condition. The method comprises analysing a nucleic acid of
XX a sample from the subject and detecting the presence of a mutation
XX (e.g. the pickwick mutation in the cardiac specific exon N2B) in the
XX titin gene, which indicates that the subject has or is at risk of
XX developing a titin-related disease. The zebrafish which has a phenotype
XX similar to mammalian heart failure is used as a model. The method is
XX useful for detecting an increased likelihood of heart disease, such as
XX heart failure, in a patient, so that appropriate intervention can be
XX instituted before any symptoms occur. The method may also be used to
XX facilitate determination of etiology of an existing heart condition,
XX such as heart failure, to identify compounds that can be used to treat
XX or prevent heart conditions, in prenatal genetic screening, e.g. to
XX identify parents who may be carriers of a recessive titin mutation.
XX Compounds identified using the methods may be used to treat patients
XX that have or are at risk of developing heart disease, e.g. heart
XX failure.
```

SQ Sequence 26926 AA;

alignment_scores:

Quality:	134.00	Length:	851
Ratio:	0.398	Gaps:	40
Percent Similarity:	39.600	Percent Identity:	18.801

alignment_block:

US-09-049-696-20 x AAU05396 ..

Align seg 1/1 to: AAU05396 from: 1 to: 26926

```
610 GTAGTAAAGAAGTGTACAGGAGGCGAGCTGTACACCAAAAGATGCCACATT 659
.....:.....:.....:.....:.....:.....:.....:.....:
6830 IleileGluArgCysGluGluGlyLysAspAsnTrpIleArgCysAsnMe 6846
660 CAAT.....:.....:.....:.....:.....:.....:.....: 688
.....:.....:.....:.....:.....:.....:.....:.....:
6846 tLysLeuValProGluLeuThrTyrLysValThrGlyLeu....GluLysG 6862
689 GATGTGAGTTTGTTCCTCAATCCCGCAGACGGAGAGGCTTCTCTATAATG 738
||.....:.....:.....:.....:.....:.....:.....:.....:
6862 lyAsnLysTyrLeuTyrArgValSerAlaGluAsnLysAlaGlyValSer 6878
739 TTGTCACACATGTT.....:.....:.....:.....:.....:.....: 767
```

6879 AspProSerGluIleLeuGlyProLeuThrAlaAspAspAlaPheValGI 6895
||||:|||||
768 ATTCTGTACAGACAAACCAACAAAGAGCTCCAAACAGCAAATC 817
|||:|||||
6895 upThrMetAspLeuSerAlaPheLysAspGly. 6906
|||:|||||
818 AAAATGCAATCTCGAAGACATGGGAAGGATCCGTGATTCTGAGGAC 867
|||||
6907LeuGluValIle. 6910
|||
868 TTAAAGAAACCACTCTATGACACACAGCCCAATCCCACTTCTC 917
|||||
6911ValProAsnPro. 6914
918 ATTGCTGCAGATTGCACAAAGATTGTGTAGTCTTGACAAATCTG 967
|||||
6915IleThrIleLeuValProSerThrGlyT 6924
|||||
968 GAAGCATGGCAGCTGGTAACCGCTCAATCGACTGAATCAAGCAGGCCAG 1017
|||
6924 yProArgProThr. 6931
|||
1018 CTTTTCCTGCTGCACAGAGTTGAGCTGGGTCCTGGGTGGATGGTGAC 1067
|||:|||||
6932 CysPheGlyAspLysValLeuGluThrGlyAspArgValLysMetLysTh 6948
|||
1068 ATTTCACAGTGTGCCATGTACAAAGTGAACCTCATACAGATAAAGCTG 1117
|||||
6948 rLeuSerAlaTyrAlaGluLeu. 6962
|||:|||||
1118 GCAGTGACAGGACACACTCGCCAAAGATTACCTGCAGCAGCTTCAGGA 1167
|||||
6962 rSerAspLysGlyIleThrLeuLysLeuGlu. 6973
|||:|||||
1168 GGGAGCTCCATCTGCAGCGGCTTCGATCGCGCATTTACTGTGATTAGGAA 1217
|||
6973 6973
1218 GAAATATCAACTGATGATGTGAATTTGCTGCTGACGGATGGGAAG 1267
|||
6973 6973
1268 ACAACACTATAAGTGGTCTTTACGAGGTCAAACAA. .AGTGGTGCC 1314
|||:|||||
6974AsnArgValLysThrIleSerGlyGlu 6982
|||
1315 ATCATCCACACAGTGCCTTTGGGGCCCTCGCAGCTCAAGAACTAGAG. 1362
|||:|||||
6983 IleAspValAsnValIleAlaArgProSerAlaProLysGluLeuLysPh 6999
|||:|||||
1363GAGCTGTCCAAATGACAGGAGGTTTACAGACATATGCTTCAGATC 1408
|||:|||||
6999 eGlyAspIleThrLysAspSerValHisLeuThrTrpGluProAsp. 7015
|||
1409 AGTTTCAGAACATGCGCTCATTTGATCTTTGGGGCCCTTTTCATCAGGA 1458
|||
7016AspAspGlySerProLeuThrGly 7024
|||
1459 AATGGAGCTGTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAC 1508
|||:|||||
7025 Tyr...ValValGluLysArgGluValSerArgLysThrTrpThrLysVa 7040
|||:|||||
1509 CTCCAGACAGCCAGTGGATGAATGGCAGAGTATGCTGGGACAGCACCG 1558
|||:|||||
7040 lMetAspPheValThrAspLeuGluPheThrVal...ProAspLeuValG 7056
|||
1559 TGGGAAGAGACCTTGTCTTTCTTATC. 1584
|||:|||||
7056 lnGlyLysGluTyrLeuPheLysValCysAlaArgAsnLysCysGlyPro 7072
|||
1585ACCTG 1589
|||

7073 GlyGluProAlaTyrValAspGluProValAsnMetSerThrProAlaTh 7089
|||:|||||
1590 GACAAGCAGCCTCCCAAA. 1608
|||:|||||
7089 rValProAspProGluAsnValLysTrpArgAspArgThrAlaAsnS 7106
|||:|||||
1609ATCCTTCTCTGGGATCCC. 1641
|||:|||||
7106 erIlePheLeuThrTrpAspProProLysAsnAspGlySerArgIle 7122
|||:|||||
1642 GGTGGCTTTGTGTGACAAA. 1662
|||:|||||
7123 LysGlyTyrIleValGluArgCysProArgGlySerAspLysTrpValAl 7139
|||:|||||
1663ACACCAAATGGCCTACCTCCCAATCCCAAG 1693
|||:|||||
7139 acGlyGluProValAlaGluThrLysMet. 7153
|||:|||||
1694 GCATTGCTAAGTTGGCCTGG. 1725
|||:|||||
7153 lLeuGluGlu...GlyLysTrpTyrAlaTyrArgValLysThrLeuAsn 7168
|||:|||||
1726 CTGCAAGCAAGCTCACAA. 1743
|||:|||||
7169 ArgGlnGlyAlaSerLysProSerArgProThrGluGluIleGlnAlaVa 7185
|||:|||||
1744ACCTTGACCCCTGA 1756
|||
7185 lAspThrGlnGluAlaProGluIlePheLeuAspValLysLeuAlaG 7202
|||
1757 CTGTCAGTCCCGTGGCTCCAATGCTACCTGCTCCCAATACAGTGACT 1806
|||:|||||
7202 lLeuThrValLysAlaGlyThrLysIleGluLeuProAlaThrValThr 7218
|||:|||||
1807 TCCAAAACGAACAGGACACCAATTCGCCAGCCTCTGTAGTTTA 1856
|||:|||||
7219 GlyLys. 7227
|||:|||||
1857 TGCAATATTCGCCAAGGAGCCTCCCAATTCACGGCCAGGTGCACAG 1906
|||:|||||
7227 pThrLys. 7240
|||:|||||
1907 CCCTGATTGAATCAGTCAATGGAAAA. .ACAGTTACCTTGGAACTACTG 1953
|||:|||||
7240 lThrIleGluAsnValProLysLysSerThrValThrIle. 7253
|||:|||||
1954 GATATGGAGCAGGTGCTGATGCTACTAGGATGACGGTGTCTACTCAAG 2003
|||:|||||
7254ValAspSerLysArgSerAsp. 7260
|||:|||||
2004 GTATTTCAAACTTATGACACGAATGGTAGATACAGTGTAAAGTGGCGG 2053
|||:|||||
7261ThrGlyThrTyrIleIleGlu. 7267
|||:|||||
2054 CTCGGGAGGAGTTAACCCAGCCAGAGGAGAGTATACCCACAGAGT 2103
|||:|||||
7268AlaValAsnValCysGlyArgAlaThrAlaValValGluVal 7281
|||:|||||
2104 GGACACTGTACATACCTGGC. 7299
|||:|||||
7282 AsnValLeuAspLysProGlyProProAlaAlaPheAspIleThrAspVa 7298
|||:|||||
2130 TGAGAAATGATGAA. .ATACAAATGGAATCCCAAGACCTGAA. 2169
|||:|||||
7298 lThrAsnGluSerCysLeuLeuThrTrpAsnProProArgAspGlyG 7315
|||:|||||
2170ATTAATAAGGATGAT 2184
|||:|||||
7315 lYserLysIleThrAsnTyrValValGluArgAlaThrAspSerGlu 7331
|||:|||||
2185 GTTCAACACAAAGAGTGTGTTTTCAGACAGACATCTCGGAGGCTCAT 2234
|||:|||||
7332 ValTrpHisLys. 7345
|||:|||||
.....LeuSerSerThrValLysAspThrAsnPh

2235 TGTGGCTTCTGAT...CTCCCAAT..... 2256
| | | | | : | | | | |
7345 eLysAlaThrLysLeuileProAsnLysGluTyrllePheArgValAlaa 7362
2257GCTCCCATACCT 2268
7362 laGluAsnMetTyrglyAlaGlyGluProValGlnAlaSerProIleThr 7378
2269 GATCTTTC.....CCACTGGCCAAATCACCAGCTGAAGGCGGA 2309
: | | | | | : | | | | |
7379 AlaLysTyrglnPheAspProGlyProThrArgLeuGluProSe 7395
2310 AATTACGGGGGAGCTCTATTAACTGACTGGACAGCTCTCTGGGGATG 2359
: : : : : | | | | |
7395 rAspIleThrLysAspAlaValThrLeuThrTrpCysGluProAspAspA 7412
2360 ATTATGACCATGGAACAGCTCACAGTATATCATTCGTAATAAGTACAAGT 2409
|| : : : : : | | : : : : :
7412 spGlyClySerProIleThrGlyTyrrpValGluArgLeuAspProAsp 7428
2410 ATCTTGATCTCAGAGACAAGTTC.....AATGAATCTCTCAAGTGAA 2453
| | | | | : | | | | |
7429ThrAspLysTrpValArgCysAsnLysMetProVally 7441
2454 TACTACTGCTCTCATCCCAAGAGCCCACTCTGAGGAAGTCTTTTGT 2503
: | | | | : : : : : : : : : : :
7441 sAspThrThrTyrrArgVallyGlyGlyLeuThrAsnLysLysLysTyrrArgp 7458
2504 TTAACACGAGAAACATTTCTTTGAAATGGCAGACATCTTTTCATTGCT 2553
| | : : : | | | | : : :
7458 heArg.....ValLeuAlaGluAsnLeuAlaGlyProGlyLysPro 7471
2554 ATTCAGGCTGTGATAGCTCATCGAATCAGAAATATCCAACATTGC 2603
: : : : : : : : : : : | | : : :
7472 SerLysSerThrGluProIleLeuileLysAspProIleAsp..... 7485
2604 ACGAGTATCTTTTATTCCTCCACAGACTCCGCCAGAGACACCTAGTC 2653
| | | | |
7486ProProTrpProGlyLysProThrV 7495
2654 CTGATGAAACGTCGTCTTGTCTTAATATTCATATCAACAGACACCAT 2703
: : : : : : : : : : : | | |
7495 aLysAspValGlyLys.....ThrSerValArgLeuAsnTrpThrLys 7509
2704 CCT 2706
| | |
7510 Pro 7510

This Page Blank (uspto)


```
150 alAlaAlaAspGlyGlnThrAlaGlyGluGlyAlaSerIleSerLeuTyr 166
435 .....AG 436
167 ThrLysProSerProValLeuAspLeuLysAlaGluTyrValGlyValTh 183
437 CTCACAAACCTTGACCTGACTGCAGTCCCGTGCCTCCCAATGCTACCC 486
183 rSerValAsnLeuThrTrpThrValAsnAspThrAlaSerAlaSerTyrT 200
487 TGCCTCCAAATTACAGTACTTCCAAAACGAACAGACACACAGCAAAATTC 536
200 hrTyrArgIleGluValThrSerAspSerIleAspSerLeu..... 214
537 CCCAGCCCTCTGGTAGTTATGCAAAATATCGCCAAGGAGCCTCCCAAT 586
215 .....ThrSerSerVa 218
587 TCTCAGGCCAGTGTCACAGCCCTGATTGAATCAGTGAATGGAACACAG 636
218 lThrMetAlaGluIleThrGlyLeuPro.....GlyThrLeuT 232
637 TTACCTTGGAACTACTGATTAATGGACAGGAGTGTGCTGCTACTAAGGAT 686
232 yrSerPheLysValPheAlaValAlaAlaAsp...AsnArgThrGluAla 247
687 GACGCTGTCTACTCAAGCTATTTCACA..... 713
248 AspGlyAlaSerIleSerLeuTyrThrLysProSerProValLeuAspLe 264
714 .....ACTTATGACACGA 726
264 ulysAlaGluTyrValGlyValThrSerValAsnLeuThrTrpThrValA 281
727 ATGGT.....AGATACAGTGTAAAGTGGGCTCTG 758
281 snGlyThrAlaLeuThrAlaTyrThrTyrArgIleGluValArgAsnAla 297
759 GGAGGAGTTAAAGCAGCCAGCAGGAGTGTATACCCAGCAGAGTGGAGC 808
298 ThrSerIleArgAsnGluThrSerAsnIleAsnLysIleGluIleThrGl 314
809 ACTGTACATACCTGGCTGGATTGAGATGATGAATACATAGGAATCCA. 857
314 yLeu...ileProGlyThrSerTyrAsnPheLysValPheAlaThrProv 330
857 ..... 857
330 alAsnAsnThrThrGluGluGluGlyLeuSerLeuAsnLeuTyrThrLys 346
858 CCAAGACCTGAAATTAAT..... 875
347 ProSerProValLeuArgValValThrGluTyrValGlyValThrSerVa 363
876 .....AAGATGATGTTCAACACACAGCAAGTGTGTTCA 909
363 lAsnLeuThrTrpMetValAspThrAlaSerAspSerTyrThrTyrA 380
910 GCAGAACATCTCTGGGAGGCTCATTTGTG.....GCTTCTGATGTC 950
380 rgIleGluValArgAsnGlySerSerIleAsnAsnLysThrSerAspIle 396
951 CCAATGCTCCCATCATCTGCTTCCCA..... 980
397 ThrAspAlaGluIleThrGlyLeuIleProGlyThrLeuTyrThrPheTh 413
980 ..... 980
413 rValPheAlaValAlaAlaAspGlyGlnThrAlaGlyGluGlyAlaSerI 430
981 .....CCTGGCCAAATCACCAGCTGAAGCGGAAAT 1013
430 leSerLeuTyrThrLysProSerProValLeuAspLeuLysAlaGluTyr 446
```

```
1014 CACGGGGCAGTCTCATTAATCTGACTTGGACAGCTCCTGGGGATGATTA 1063
447 ValGlyValThrSerValAsnLeuThrTrpThrVal..... 458
1064 TGACCATGGACAGCTCACAAAGTATATCATTCGAATAAGTACAAAGTATTC 1113
459 .AsnAspThrAlaSerAlaSerTyrThrTyrArgIleGluValThr.... 473
1114 TTGATCTVCAGAGACAAGTTCAATGAATCTCTTCAAGTGAATACT..... 1157
474 .....SerAspSerSerIleAspSerLeuThrSerAsnValThrMet 487
1158 .....ACTGCTCTCATCCAAAGAGCAACTCTCGAGGAAGTCTT 1198
488 AlaGluIleThrGlyLeuIleProGlyThrLeuTyrAsnPheThrValPh 504
1199 TTGTGTTAAACAGAAACATTACTTTTGAATAATGGACAGATCTTTTCA 1248
504 eAlaValAlaAlaAspAsnArgThrGluAlaAspGlyAlaPheThrSerL 521
1249 TTGCTATTACAGGCTGTGATAAGTCTGATGAAATCAGAAATATCAAC 1298
521 eutyThrLysProThrProValThrAspLeuLysAlaGlu...HisGly 536
1299 ATTGCACGAGTATCTTTG 1316
537 ValThrSerValSerLeu 542
seq_name: sp_rodent:O55005
seq_documentation_block:
ID O55005 PRELIMINARY; PRT; 1651 AA.
AC O55005;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE TRANSMEMBRANE RECEPTOR ROBO1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPINAL CORD;
RX MEDLINE=98117249; PubMed=9458045;
RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA Goodman C.S., Tear G.;
RT "Roundabout controls axon crossing of the CNS midline and defines a
RT novel subfamily of evolutionarily conserved guidance receptors.";
RL Cell 92:205-215(1998).
CC -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL; AF041082; AAC39960.1; -.
DR HSSP; P56276; 1TLK.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00041; In3; 3.
DR Pfam; PF00047; Ig; 5.
DR SMART; SM00060; FN3; 3.
DR SMART; SM00408; IGC2; 5.
KW Transmembrane.
SQ SEQUENCE 1651 AA; 180746 MW; FA2452DD46E186B7 CRC64;
```

```
alignment_scores:
  Quality: 135.00      Length: 467
  Ratio: 0.619        Gaps: 23
Percent Similarity: 46.681 Percent Identity: 21.413
alignment_block:
US-09-049-696-19 x O55005 ..
```


Align seg 1/1 to: 055005 from: 1 to: 1651

```

228 ATGAATGGCACAGTGTATGCTGGACACGACCGTGGAAAGGACACTTTGTT 277
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
462 ValAsnGlnThrValAlaValAspGlyThrLeu.....ThrLeuSe 475
278 TCTTATACCTCGGACACGACGACCTCCCAAAATCCTTCTCTGGATCCCA 327
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
475 rCysValAla...ThrGlySerProValProThrIleLeuTrp..... 488
328 GTGGACAGAACAGGTGGCTTTGTAGTGACACAAACACCAAAATGGCC 377
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
489 .....ArgLysAspGlyValLeuValSerThrGlnAspSerArgIleLys 503
378 TACCTCCAAATCCCGGCATT.....GCTAAGGTGGC... 410
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
504 GlnLeuGlu...SerGlyValLeuGlnIleArgTyrAlaLysLeuGlyAs 519
411 .ACTTGGAAATACAGTCTCAAGCAGCTCACAACCTTGACCTGACATG 459
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
519 pThrGlyArgTyrThrCysThrAlaSerThrProSerGlyGluAlaThrT 536
460 TCAGTCC..... 467
536 rPserAlaTyrIleGluValGlnGluPheGlyValProValGlnProPro 552
468 CGTGGGTCCAATGCTACCTGCTCCCAATTACAGTACGATTCACAAACGAA 517
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
553 ArgProThrAspProAsnLeuIleProSerAlaProSerLysProGluVa 569
518 CAAGGACACCAACCAATCCCGCAGCTCTGCTAGTTATGCAAAATATTC 567
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
569 lThrAspValSerLysAsnThrValThrLeuLeuTrpGlnProAsnLeuA 586
568 GCCAAGGACCTCCCA.....ATTCTCAGGCGC..... 596
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
586 snSerGlyAlaThrProThrSerTyrIleIleGluAlaPheSerHisAla 602
597 .....AGTGTACAGCCCTGATGATGATGATGATGATGATGATGATGAT 637
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
603 SerGlySerSerTyrPThrValAlaGluAsnValLysThrGluThrPh 619
638 TACCTTGGNA.....CTACTGGATATG 660
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
619 eAlaIleLysGlyLeuLysProAsnAlaIleTyrLeuPheLeuValArgA 636
661 GAGCAGGTGCTGATGCTACTAAGGATGACGGTCTCTACTCAAGGTATTTTC 710
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
636 laAlaAsnAlaTyrGlyIleSerAspProSerGlnIleSerAspProVal 652
711 ACAACTTATGAC.....ACGAATGTAGATACAGTGTAAAGT 748
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
653 LysThrGlnAspValProProThrThrGlnGlyValAspHisLysGlnVa 669
749 G...CGGGCTCTGGAGGAGTT..... 767
669 lGlnArgGluLeuGlyAsnValValLeuHisLeuHisAsnProThrIleL 686
768 .....AACGACGACGACGAGAGTATACCCACGACAGT... 803
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
686 euSerSerSerValGluValHisThrThrValAspGlnGlnSerGln 702
804 .....GGACACTGTACATACCTGGCTGGATGACAGATGA 838
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
703 TyrIleGlnGlyTyrLysIleLeuTyrArgProSerGlyAlaSerHisGl 719
839 TCAATACATCG.....AATCCA..... 857
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
719 yGluSerGluTrpLeuValPheGluValArgThrProThrLysAsnSerV 736
858 .....CCAACA 863
      |||

```

```

736 alValIleProAspLeuArgLysGlyValAsnTyrGluIleLysAlaArg 752
864 CCTGAATTAATAAGGATGATGTTCAACACAAGCAAGTGTGTTTTCAGCAG 913
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
753 ProPhePheAsnGluPheGlnGlyAlaAspSerGluIleLysPheAlaLys 769
914 ACATCTCTCGGAGGCTCATTTGTGGCTTCTGATGTCCTCCCAATGCTCCCA 963
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
769 sThr.....LeuGluGluArgProSerAla.... 777
964 TACCTGATCTCTTCCACCTGCGCAATCACCAGCCTGAAGCGGAAAT 1013
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
778 .....ProProArgSerValThrValSerLysAsnAspGly 789
1014 CACGGGGCAGTCTCATTAATCTGACTTGGACAGCTCTCTGGGATGATTA 1063
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
790 AsnGlyThrAlaIleLeu...ValThrTrpGlnProProGluAspTh 805
1064 TCACCATGGAACAGCTCACAGTATATCATTCGAATAAGTACAAATATTC 1113
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
805 rGlnAsnGlyMetValGlnGluTyrLysValTrpCysLeuGlyAsn.... 820
1114 TTGATCTCAGACAGCAAGTTCAATGAATCTTCAAGTGAATACTACTGCT 1163
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
821 ..GluThrArgTyrHisIleAsnLysThrValAspGlySerThrPheSer 836
1164 .....CTCATCCCAAGAGCAACTCTCAGGAAGTC... 1196
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
837 ValValIleProPheLeuValProGlyIleArgTyrSerValGluValAl 853
1197 .....TTTTTGTTTAAACCAGAAACATTA 1221
853 aAlaSerThrGlyAlaGlyProGlyValLysSerGluProGlnPheIleG 870
1223 CTTTGAATAATGGCAGACATCTTTTCATTGCTATTCCAGGCTGTGATAAG 1271
870 lLeuAspSerHisGlyAsn.....ProValSerProGluAspGln 883
1272 GTCGATCTGAATCAGAAATATCCAACTTGCACGAGTATCTTTGTTTAT 1321
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
884 ValSerLeuAlaGlnGlnIleSerAspValValLysGlnProAlaPheI 900
1322 T 1322
900 e 900
seq_name: sp_invertebrate:Q9V3J6
seq_documentation_block:
ID Q9V3J6 PRELIMINARY; PRT; 721 AA.
AC Q9V3J6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG5264 PROTEIN.
GN BEST OR CG6264.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Hallow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

```

```

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieux E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle K., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam K.A.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.V., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., McCarthy C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.R., Puri V., Reese M.G.,
RA Reineck K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J.H., Yen R.-F., Zaveri F.N., Zhong W., Zhou X., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.W., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RA "The genome sequence of Drosophila melanogaster.";
RA Science 287:2185-2195(2000).
RL [2]
RN SEQUENCE FROM N.A.
RP Tavanali B.C., Pappu K., Mardon G.:
RT "Submitted (DEC-1999) to the EMBL/GenBank/DBSJ databases:
RL EMBL; AE003686; AAF54503.1; -
DR EMBL; AF218817; AAF32327.1; -
DR FlyBase; FBgn0040238; best.
DR InterPro; IPR000615; Worm.fam.8.
DR Pfam; PF01062; Worm.family.8; 1.
DR ProDom; PD002802; Worm.fam.8; 1.
DR PROSITE; PS00000; Worm.fam.8; 1.
SQ SEQUENCE 721 AA; 79642 MW; B37CD96DE2F15B15 CRC64;

alignment_scores:
Quality: 134.50 Length: 545
Ratio: 0.538 Gaps: 31
Percent Similarity: 45.872 Percent Identity: 21.651

alignment_block:
US-09-049-696-19 x Q9V336 ..

Align seg 1/1 to: Q9V336 from: 1 to: 721

57 CTAGAGGAGCGTGTCCAAATATGACAGAGGAGGTTCACAGACATATGCTTCAGA 106
215 ILEAspEluteLeuAsnLysPheArgGly..... 223
107 TCAAGTTCAGAACAAATGGCTCATTCATGCTTTGGGGCCCTTTCATCAG 156
224 .....GlnCysGlyLeuLeuLeuSerTyrAspThrIleSerValP 237
157 GAAATGGAGCGTGTCTCTCCAGCGCTCCATCCAGCTTGACAGATGAAGGATTAA 206
237 roLeuValTyrThrGlnValValThrLeuAlaValTyrSerTyrPheLeu 253
207 ACC...CTCCAGACACGCGATGGATGAATCGACAGCTGATCGTGGACAG 253
254 ThrCysCysMetGlyGlnGlnTrpThrAspGlyLysValValGlyAsnTh 270
254 CACC...GTGGGAAAGACGACTTTGTTTCTTATACCTCGGACACGACCC 300
270 rThrTvrLeuAsnLysValAspLeuTyrPheProValPheThrThrLeu. 286

```

```
979 CACCTGGCCAAATACCGACCTG.....AAGCGGAAATTCACGGGGC 1022
||||| : : : |||
563 roProGlyValAlaLeuSerThrAlaProAlaGlyGly 579
||||| : : : |||
1023 ACT.....CTCATTAATCTGACTTGGACAGCTCTGGGATGATTATGA 1066
||||| : : : |||
580 SerProLeuLeuSerAsnAlaThrAlaProSerProValG1 596
||||| : : : |||
1067 CCGTGAACAGCTCACAGTATATCATTCGAATAAGTACAGTATCTTG 1116
||||| : : : |||
596 yGluSer.....SerLysSerLeuTyra 604
||||| : : : |||
1117 ATCTCAGACAGAAGTTCAATGAATCTCTCAAGTGAATCTACTGCTC 1166
||||| : : : |||
604 sp.....
1167 ATCCCAAGGAAGCAACTCTCAGGAAGCTTTTGTGTTTAAACAGAAAA 1216
||||| : : : |||
605 ...ProGlnLysGlyAlaSerArgGluThrVal.....GluSe 616
1217 CATTACTTTTGAATGGACAGATCTTTTC.....ATTGCTATTTCAGG 1260
||||| : : : |||
616 rMetAspLeuArgSerSerThrAspLeuLeuGlyAspAlaAlaValGlnP 633
||||| : : : |||
1261 CTGTTGATAAGTCGAT.....CTGAATACAGAA..... 1289
||||| : : : |||
633 roGluAspGluGlyAspAspPheAspLysLeuLysAlaGluArgGluLys 649
||||| : : : |||
1290 .....ATATCCAACATTCACGAGTATCTTTGTTTAT 1321
||||| : : : |||
650 GluLysLeuMetArgGlnGlnLysAsnLeuAlaArg..... 661
1322 TCCTCCACAGACTCGCCAGACACCTAGTCTCTGATGAACGTCTGCTC 1371
||||| : : : |||
662 .....ThrIleSerThrAlaProGlyMetGluAlaThrAlaValP 675
||||| : : : |||
1372 CT...TGTCCTAATATTCATATCAACAGCACCATT 1403
||||| : : : |||
675 roMetValProMetValProValAsnValAlaVal 686
||||| : : : |||
seq_name: sp_vertibrate:Q90783
seq_documentation_block:
ID: Q90783 PRELIMINARY; PRT: 1173 AA.
AC Q90783;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ZINC FINGER PROTEIN PROCKR2 (FRAGMENT).
GN CKR2.
OS Gallus gallus (Chicken).
OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW CELLS;
RA MEDLINE=95047430; PubMed=7958949;
RA Schuetz B., Niessing J.;
RT "Cloning and structure of a chicken zinc finger cDNA: restricted
RT expression in developing neural crest cells.";
RL Gene 148:227-284(1994).
DR EMBL; X56805; CAA40140.1; -.
DR HSSP; P08047; 1SP2.
DR InterPro; IPR000822; Znf-C2H2.
DR Pfam; PF000096; Zf-C2H2; 16.
DR SMART; SM00355; Znf-C2H2; 16.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 15.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 15.
KW DNA-binding; Metal-binding; Zinc-finger.
FT NON_TER
FT 1
SQ SEQUENCE 1173 AA; 127327 MW; 493FF4C07300FA10 CRC64;
```

```
alignment_scores:
Quality: 134.00 Length: 535
Ratio: 0.565 Gaps: 21
Percent Similarity: 44.299 Percent Identity: 19.626
alignment_block:
US-09-049-696-19 x Q90783 ..
Align: seg 1/1 to: Q90783 from: 1 to: 1173
159 AATGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGGA..... 203
||||| ||| ||| : : : : : |||
519 AspGlyProValLeuGlnAlaLeuSerLeuGluValThrAlaProAs 535
204 .....TTAACCTCCAGAACAGCGATGATGAGTGGCACAG 240
||||| ||| ||| : : : : : |||
535 palahisThrPheLeuLeuGlnThrProGlnGlyLeuGln.....L 550
241 TGATCGTGACAGCACCGGTGGAAAGGACACTTTGTTCTTATCCTCG 290
||||| : : : : : |||
550 euileProSerProProAlaProGlnLysLeuLeuLeuProThr 566
291 ACAACGACGCTCC.....CA 307
||||| : : : : : |||
567 AlaProGlnProProLysHisGlnGlyGluSerProThrProGlyG 583
308 AATCTCTCTCGGATCCCGTGA.....CAGA 336
||||| ||| ||| : : : : : |||
583 nSerLeuLeuValProSerThrGlyThrThrLeuProThrLeuArgL 600
337 AGCAAGTGGCTTTGTAGTGACAAACCAAAATGCGCTACCTCCAA 386
||||| : : : : : |||
600 euGlnAlaValThrAlaValProGlnGlyThrGlyThrGlyLeuVal 616
387 ATCCAGGCGATT..... 398
||||| : : : : : |||
617 LeuGlnGlyLeuProGluGlnProLeuHisProAlaGlyLeProHisG 633
399 ....GCTAAGTTGGCATTGGAAATACAGTCTGCAAGCAAGCTCACA 444
||||| ||| ||| : : : : : |||
633 yGlnAlaAlaValGluGlyAlaAlaValArgLeuGlnAlaAsnGluVal 650
445 CC...TTGACCTGACTGTACGTCCTCCGTCGTCCTACCTCGCT 491
||||| : : : : : |||
650 hrAsnValGlnLeuGlnAlaLeuProGlnProThrAspValThrSerI 666
492 CCAATTACAGTGACTCCAAACGACACAGGACACCAAAATTCGCCAG 541
||||| : : : : : |||
667 ProLeuGlnAlaThrAspMetThrAsnValGlnLeuGlnAlaLeuPro 683
542 CCCTCTGGTAGTTTATGCAATATTCGCCAAGGAGCTCCCAATTTCTCA 591
||||| ||| ||| : : : : : |||
683 nProThrAspVal...ThrAsnIleGlnLeuGlnAlaThrGluValThr 699
592 GGGCCAGTGTACAGCCCTGATGATGATGATGATGATGATGATGATGAT 641
||||| : : : : : |||
699 snValGlnLeuGlnAlaLeuLeuGlnProThrAspValThrAsnIleG 715
642 TTGGAACTACTGGATAATGGACAGGCTGCTGCTACTAAGGATGACGG 691
||||| : : : : : |||
716 MetGlnAlaThrGluValAla..... 722
692 TGTCTACTCAAGGTATTTCACTTATGACACGAATGTTAGATACAGTG 741
723 .....AsnV 724
742 TAAAGTGGCGGCTCTG..... 758
||||| : : : : : |||
724 aiGlnLeuGlnAlaLeuSerGlnProProAspValThrAsnIleGluLe 740
759 .....GGAGGAGTTAACGACGACGACGACGAGTATACCCAGCAGAG 802
```


[illegible]

```
seq_documentation_block:
ID 089026 PRELIMINARY; PRT; 1612 AA.
```

AC	089026;
AD	01-NOV-1998 (TReMBLrel. 08, Created)
AE	01-NOV-1998 (TReMBLrel. 08, Last sequence update)
AF	01-JUN-2001 (TReMBLrel. 17, Last annotation update)
AG	DUT11 PROTEIN.
AH	RGE01 OR DUT11.
AI	Mus musculus (Mouse).
AJ	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AK	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AL	[1]
AM	SEQUENCE FROM N.A.
AN	TISSUE=BRAIN;
AO	Wu M.C., Lowe N., Fordham R., Rabbitts P.;
AP	"the mouse homologue of human DUT11/H-robo1 gene: protein sequence and
AQ	chromosomal location.";
AR	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
AS	-1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
AT	DOMAIN.

DR	SMART; SM00410; IG_like; 4.	5F2988C544796B4B CRC64;
SQ	SEQUENCE 1612 AA; 176406 MW;	

```

alignment_scores:      Quality: 128.00      Length: 467
                       Ratio: 0.593        Gaps: 22
Percent Similarity: 46.253 Percent Identity: 20.771

alignment_block:
US-09-049-696-19 x 089026 ..

Align seq 1/1 to: 089026 from: 1 to: 1612

228  APTGAATGGCAGTGAATCGTGTGACACACCGCTGGGAAGACACATTGTGT 277
      ::::: ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
423  ValAsnGlnThrValAlaValAspGlyThrLeu.....IleLeu 435

278  TCTTATCACTCGACACGACGCTCCCAAAATCCTTCTCTGGGATCCCA 327
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
435  uSerCysValAlaThrGlySerProAlaProThrIleLeuTpr..... 449

328  GTGGACAGAAAGTGGCTTTGTAGTGGACAAACACCAAAATGGCC 377
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
450  ....ArgLysAspGlyValLeuValSerThrGlnAspSerArgIleLys 464

378  TACCTCCAAATCCAGGCATT.....GCTAAGGTTGGC... 410
      ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
465  GlnLeuGlu...SerGlyValLeuGlnIleArgTyrAlaLysLeuGlyAs 480

411  .ACTTGGAAATACAGTCTGCAACGACAGCTCACAAACCTTCACCTGAC 459
      ||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
480  pThrGlyArgTyrThrCysThrAlaSerThrProSerGlyGluAlaThr 497

460  TCAGGTCC..... 467
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
497  rPserAlaTyrIleGluValGlnGluPheGlyValProValGlnProPro 513

468  CGTGGTCCAAATGCTACCTGCCTCCCAATTACAGTGACTTCCAAAACGAA 517
      ||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
514  ArgProThrAspProAsnLeuIleProSerAlaProSerLysProGluVa 530

518  CAAGGCACACAGCAAAATTCGCCAGCCCTCGGTAGTTATGCAAAATATC 567
      ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
530  lThrAspValSerLysAsnThrValThrLeuSerTprGlnProAsnLeuA 547

568  GCCAAGAGGCTCCCA.....ATTCTCAGGGCC..... 596
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
547  snSerGlyAlaThrProThrSerTyrIleIleGluAlaPheSerHisAla 563

597  ....AGTGTCAAGGCTGATTGATTCAGTCAAGTGAAGGAAAAACAGT 637
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
564  SerGlySerSerTprGlnThrAlaAlaGluAsnValLysThrGluThrPh 580

638  TACCTTGGAA.....CTCTGGGATATG 660
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
580  eAlaIleLysGlyLeuLysProAsnAlaIleTyrLeuPheLeuValArgA 597

661  GAGCAGGTGCTGATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTC 710
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
597  laAlaAsnAlaTyrGlyIleSerAspProSerGlnIleSerAspProVal 613

711  ACAACTTTATGAC.....ACGAATGCTAGATACAGTGTAAAAAGT 748
      ||| ||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
614  LysThrGlnAspValProProThrSerGlnGlyValAspHisLysGlnVa 630

749  G...CGGGCTCTGGGAGGAGTT..... 767
      ||| ||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
630  lGlnArgGluLeuGlyAsnValValLeuHisLeuHisAsnProThrIleL 647

768  ....AACGCAGCCACGAGAGTATACCCACGACAGAGT... 803
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
647  euSerSerSerValGluValHisThrThrValAspGlnGlnSerGln 663

804  ....GGAGCACTGTACATPACTCGCTGGATTGACAGATGA 838
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
664  TvrIleGlnGlyTyrLysIleLeuTyrArqProSerGlyValAspHisG 680
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

```

```

839 TGAATCAATG...AATCCA... 857
    ||| :|||
680 yGluSerGluTrpLeuValPheGluValArgThrProThrLysAsnServ 697
    858 .....CCAAAGA 863
        |||
697 alValIleProAspLeuArgLysGlyValAsnTyrgLuIleLysAlaArg 713
    864 CCTGAATTAATAAGGATGTTCAACACACACCAAGACTGTCTTTCAGCAG 913
        ||| :||| :||| :||| :||| :||| :||| :|||
714 ProPheAsnGluPheGlnGlyAlaAspSerGluIleLysPheAlaIalY 730
    914 AACATCTCGGGAGGCTCATTTCTGGCTCTCTGATGTGCCAAATGCTCCCA 963
        ||| :||| :||| :||| :||| :||| :||| :|||
730 sThr.....LeuGluGluAlaProSerAla.... 738
    964 TACCTGATCTCTCCACCTGGCCAAATCACCGACCTGAAGCGGNAATT 1013
        ||| :||| :||| :||| :||| :||| :||| :|||
739 .....ProArgSerValThrValSerLysAsnAspGly 750
    1014 CACGGGGCAGCTCACTCAATTAATCTGACTTGACACGCTCTGGGATGATTA 1063
        ||| :||| :||| :||| :||| :||| :||| :|||
751 AsnGlyThrAlaIleLeu...valThrTrpGlnProProGluAspTh 766
    1064 TGACCATGGACAGCTCAACAAGTATATCATTCGANAAGTACAAGTATTC 1113
        ||| :||| :||| :||| :||| :||| :||| :|||
766 rGlnAsnGlyMetValGlnGluTyrlYsValTrpCySLeuGlyAsn.... 781
    1114 TTGATCTCAGACACAGTTCATCAATCTCTCAAGTGAATACT..... 1157
        ||| :||| :||| :||| :||| :||| :||| :|||
782 ..GluThrLysTyzHisIleasnLysThrValaspGlySerThrPheSer 797
    1158 .....ACTGCTCTCATCCCAAAGAACCCAACTCTCAGGAAGTC... 1196
        ||| :||| :||| :||| :||| :||| :||| :|||
798 ValValIleProSerLeuValProGlyIleArgTyrservalGluValAl 814
    1197 .....TTTTGTTAAACACGAAAAACATTA 1221
        ||| :||| :||| :||| :||| :||| :||| :|||
814 aaLaSerThrGlyAlaGlyProGlyValLysSerGluProGlnPheileG 831
    1222 CTTTTGAAAATGGCAGACATCTTTTCATGTGATTCACGGCTGTGTGAAG 1271
        ||| :||| :||| :||| :||| :||| :||| :|||
831 lnLeuAspSerHisGlyAsn.....ProValSerProGluAspGln 844
    1272 GTCGATCTGAATCAGAAATATCCAAACATTCACAGCATCTTTGTGTTAT 1321
        ||| :||| :||| :||| :||| :||| :||| :|||
845 ValserLeuAlagInIleaserAspValValArgGlnProAlaPheIl 861
    1322 T 1322
        |
        861 e 861
seq_name: sp_human:Q9H3V5
seq_documentation_block:
ID Q9H3V5 PRELIMINARY; PRT; 4370 AA.
AC Q9H3V5;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HEPARAN SULFATE PROTEOGLYCAN PERLECAN (FRAGMENT).
GN GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleo
QC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Genoscope;
RL Submitted (OCT-2000) to the EMBL/Genbank/DBSJ databases.
[2]
RP SEQUENCE FROM N.A.
```



```

1898 rgLeuProAlaValGluProThrAspGlnAlaGlnTyrLeuCysArgAla 1914
1068 CATGGAACAGCTCAACAGTAGTATCATTCGAAATAAGTACAAGTATTTGTA 1117
1915 HisSerAlaGlyGlnGlnValAlaArg.....AlaValLeuHi 1928
1118 TCTCAGACACAAGTTCATGAATCTCTCAAGTGAAT..... 1154
1928 sValHisGlyGlyGlyProArgValGlnValSerProGluArgThrG 1945
1155 .....ACTACTGCTCTC..... 1166
1945 lnValHisAlaGlyArgThrValArgLeuTyrCysArgAlaAlaGlyVal 1961
1167 .....ATCCCAAGGA 1177
1962 ProSerAlaThrIleThrTyrArgLysGluGlySerLeuProGlu 1978
1178 ACCCAACTCTGAGAA.....GTCCTTTTGTAAACCCAGAAACA 1218
1978 nAlaArgSerGluArgThrAspIleAlaThrLeuLeuIleProAlaIle 1995
1219 TTACTTTTGAATGGCACAGATCTTTTCATTGCTATTTCAGGCTGTGAT 1268
1995 hrThrAlaAspAlaGlyPheTyrLeuCysValAlaThrSerProAlaGly 2011
1269 AAGTGTGATCTGAATCAGAAATATCCACATTCACAGAGTATCTTTGTT 1318
2012 ThrAlaGlnAlaArgIleGlnValVal...ValLeuSerAlaSerAspAl 2027
1319 TATTCCTCCACAGACTCCGCCAGAGACACTAGTCTCTGATCAACGCTG 1368
2027 aSerProProValLysIleGluSerSerProSerValThrGluG 2044
1369 CTCCTTGCTCAATATCATATCAACAGCACATTCCTGGGCATTACATT 1418
2044 ly.....GlnThrLeuAspLeuAsnCysValValAlaGlySerAlaHis 2058
1419 TTAATAATATGTGG 1433
2059 AlaGlnValThrTrp 2063

seq_name: sp_bacteria:Q9F920

seq_documentation_block:
ID Q9F920 PRELIMINARY; PRT; 973 AA.
AC Q9F920;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE SEPC.
GN SEPC.
OS Serratia entomophila.
OG Plasmid pADAP.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=42906;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ALMO2;
RX MEDLINE=20416224; PubMed=10960097;
RA Hurst M.R., Glare T.R., Jackson T.A., Ronson C.W.;
RT "Plasmid-located pathogenicity determinants of Serratia entomophila,
RT the causal agent of amber disease of grass grub, show similarity to
RT the insecticidal toxins of Photobabidus luminescens.";
RL J. Bacteriol. 182:5127-5138(2000).
DR ENBL; AFI35182; AAC09644.1;
KW Plasmid.
SQ SEQUENCE 973 AA; 107028 MW; E6F60CE457EA2A3C CRC64;

alignment_scores:

```

```

Quality: 126.50 Length: 465
Ratio: 0.623 Gaps: 23
Percent Similarity: 43.656 Percent Identity: 19.785

alignment_block:
US-09-049-696-19:x Q9F920 ..
Align seg 1/1 to: Q9F920 from: 1 to: 973

162 GGAGCTGTCTCTCAGCGCTCC...ATCCAGCTTGAGAGTAGGATTAAC 208
163 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 GlySerLeuSerGlnSerAlaAspProArgLeuHisAlaAlaGlyLeuTh 66
209 CCTCCAGAACCCAGCTGGATGAAT.....GGCAGCTGATCGTG 249
210 ||||| ||||| ||||| ||||| ||||| ||||| |||||
211 r.....AsnPheThrTyrLeuAsnSerLeuThrGlyThrValLeuGlns 81
250 ACAGCAGCTGGGAAAGACACTTTTTCATTATCATCTGGACACACGAG 299
251 ||||| ||||| ||||| ||||| ||||| ||||| |||||
252 ervAlSerAlaAspAlaGlyThr..... 88
300 CCTCCCAAAATCTTCTCTGGGATCCAGTGACAGAGCAAGGTGGCTT 349
301 ||||| ||||| ||||| ||||| ||||| ||||| |||||
302 SerLeuGluLeuSerAspAlaAlaGlyArg..... 98
350 TGTACTGGACAAAACACCAAAATCGCTACCTCCAAATCCCGGCATT. 398
351 ||||| ||||| ||||| ||||| ||||| ||||| |||||
352 AlaPheLeuAlaValThrGlyAlaG 107
399 .....GCTAAGGTTGGCAATACAGTCTGCAAGCAAGC 437
400 ||||| ||||| ||||| ||||| ||||| ||||| |||||
401 lyThrGluAspAlaValThrArgThrTrpGlnTyr..... 118
433 TCACAAACCTTGACCTGACTGTCCGCTCGCTGCTCAATGCTACCT 487
434 ||||| ||||| ||||| ||||| ||||| ||||| |||||
435 uProGlyArgProLeuSerIleThrGluGlnValThrGlyGluAlaG 140
532 AATTCCCGCCGCTCTGGTAGTTTATGCAATATTGGCCAAAGAGAGCTCC 581
533 ||||| ||||| ||||| ||||| ||||| ||||| |||||
534 lleThrGluArgPheValTyrAlaGlyAsn..... 150
582 CCAATTCTCAGGCGCAGTGTACAGCCCTGATTGAATCAGTGAATGGAA 631
583 ||||| ||||| ||||| ||||| ||||| ||||| |||||
632 AACAGTTACCTTGAACACTACTGGATAATGGAGAGGTGCTGATGCTACTA 681
633 ||||| ||||| ||||| ||||| ||||| ||||| |||||
634 ThrAspAlaGluL 155
682 AGGATGACGGTGTCTACTCAAGGTATTTACAACTTATGACAGCAATGGT 731
683 ||||| ||||| ||||| ||||| ||||| ||||| |||||
684 ysIleLeuAsnLeuAlaGlyGlnCysValSerHisTyrAspThrAlaGly 171
732 AGATACAGTGTAAAGTCCGGGCTCTGGGAGAGTTAAC....GCAGCCAG 778
733 ||||| ||||| ||||| ||||| ||||| ||||| |||||
734 LeuValGlnThrAspSerIleAlaLeuSerGlyValProLeuAlaValTh 188
779 ACGGAGAGTGATACCCAGCAGAGTGAGGAGCACTGTACATACCTGGCTGGA 828
780 ||||| ||||| ||||| ||||| ||||| ||||| |||||
781 rArgGlnLeuLeuProAspAlaAlaGlyAla.....AsnTrpM 201
823 TTGAGAATGATGAATACAATGAATGGAATCCACCAAGACCTGAAATTAATAAG 878
824 ||||| ||||| ||||| ||||| ||||| ||||| |||||
825 etGlyGluAspAlaSerAlaTrpAsn..... 209
879 GATGATGTTCAACACAGCAAGTGTGTTTTCAGCAGACATCTCTCGGGAGG 928
880 ||||| ||||| ||||| ||||| ||||| ||||| |||||
928 ...AspLeuLeuAspGlyGluThrPhePheThrGlnThrHis..... 222

```



```

929 CTCATTTGTGGCTTCGTGATGCCAAATGCTCCATACCTGATCTCTCC 978
223 .....AlaAspAlaThrGlyAlaValLeu..... 230
979 CACCTGGCCAAATCACCGACCTGAAGCGGAAATTCAC..... 1016
231 .....SerIleThrAspAlaLysGlyAsnLeuGlnArgValAlaTyr 244
1017 .....GGGGCGAGT..... 1025
245 AspValAlaGlyLeuLeuSerGlySerTrpLeuThrLeuLysAspGlyTh 261
1026 .....CTCATTAATCTGACTGGACAGCTCCTGGG..... 1055
261 rGluGlnValIleValAlaSerLeuThrTyrSerAlaAlaGlyLysLysL 278
1056 .....GATGATTATGACCATGGAACAGCTCACAGTATATCATTCGAATA 1100
278 euArgGluGluHisGlyAsnGlyValValThrSerTyrIleTyrGluPro 294
1101 AGTACAAGTATTCTT..... 1115
295 GluThrGlnArgLeuThrGlyIleLysThrGluArgProSerGlyHisVa 311
1116 .....GATCTCAGA.....GACAAGTTCA 1134
311 lAlaGlyAlaLysValLeuGlnAspLeuArgTyrThrTyrAspProValG 328
1135 ATGAATCTCTCAAGTGAATACTACTCTCTCTCATCCCAAGGAAGCAAC 1184
328 lYasnValLeuSerValAsnAsnAspAla..... 337
1185 TCTGAGCAAGCTTTTGTGTTTAAACAGAAACATTTACTTTTGAATGG 1234
338 ...GluGluThrArgPheTrpArgAsnGlnLysValValProGluAsnTh 353
1235 C.....ACAGATCTTTTCATTGCTATTCTAGGCTGTTGTGATAAGTGC 1275
353 rTyrIleTyrAspSerLeuTyrGlnLeuValSerAlaThrGlyArg.... 368
1276 ATCTGAAATCAGAAATATCCAAATTCACGAGTATCTTTGTTTATTCCT 1325
369 .....GluMetAlaAsnAlaGlyGlnGlnGlyAsnAspLeuPro 381
1326 CCACAGACTCGGCAGACAGACACCTAGTCTCTGATGAAACGCTGCT 1370
382 SerAlaThr.....AlaProLeuProThrAspSerSerAla 393

```

seq_name: sp_bacteria:031000

```

seq_documentation_block:
ID 031000 PRELIMINARY; PRT: 935 AA.
AC 031000;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE INTIMIN.
GN EAEA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98187918; PubMed=9529069;
RA Voss E., Paton A.W., Manning P.A., Paton J.C.;
RT "Molecular analysis of Shiga toxinigenic Escherichia coli O111:H-
RT proteins which react with sera from patients with hemolytic-uremic
RT syndrome.";
RL Infect. Immun. 66:1467-1472(1998).
DR EMBL; AF025311; AAC69247.1;
DR InterPro; IPR001691; GLN_synth.
DR InterPro; IPR002482; LysM.

```

```

DR InterPro; IPR003343; Big_2.
DR InterPro; IPR003344; Big_1.
DR InterPro; IPR003355; Intimin.
DR Pfam; PF01476; LysM; 1.
DR Pfam; PF02368; Big_2; 1.
DR Pfam; PF02369; Big_1; 2.
DR PRINTS; PR01369; INTIMIN.
DR PROSITE; PS00180; GLNA_1; UNKNOWN_1.
DR SMART; SM00257; LysM; 1.
SQ SEQUENCE 935 AA; 101570 MW; 406E79CDC07DEB11 CRC64;

alignment_scores:
Quality: 124.00 Length: 498
Ratio: 0.496 Gaps: 25
Percent Similarity: 50.201 Percent Identity: 21.486

alignment_block:
US-09-049-696-19 x 031000
Align seg 1/1 to: 031000 from: 1 to: 935
3 CAAAGTGGTCCCATCCACAGCTCGCTTGGGGCCCTCTGCAGCTCA 52
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
495 GlnGlyGlnIleGlnHisSer.....GlySerGlnSerAlaG 508
53 AGAAGTACAGGAGCTG.....TCCAAAATGACAGGAGTTTACAGACAT 96
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
508 nAspTyrGlnAlaIleLeuProAlaTyrValGlnGlyGlySerAsnIleT 525
57 ATGCTTCAGATCAAGTTCACAACATGGCTCATTTGATGCTTTTGGGGCC 146
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
525 yrlLys.....ValThrAlaArgAlaTyr 532
147 CTTTCATCAGGAATGAGCTGCTCTCAGCGCTCCATCCAGCTTTGAGAG 196
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
533 AspArgAsnGlyAsn.....SerSerAsnAsnValGlnLeuThrI 546
197 TAAGGATTAACCTCCAGAACAGCCAGTGGATGAATGGACAGCTGATCG 246
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
546 e.....ThrValLeuSerAsnGlyGlnValValAspGlnValGlyValT 561
247 TGGACAGCACCGTGGGAAAG.....GACACTTTGCTTTCTT 281
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
561 hrAspPheThrAlaAspLysThrSerAlaLysAlaAspGlyThrGluAla 577
282 ATCAGCTGGACAAAGCAGCGCTCCCAAAATCCTCTCTGGGATCCCACTGG 331
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
578 IleThrTyrThrAla.....ThrVal 584
332 ACAGAACAGCTGGCTTTCTAGTGGACAAAACACCAAAATGGCCTACC 381
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
584 lLysLysAsnGly.....ValThrGlnAlaAsnValProValSerPheA 599
382 TCCAAATCCCAAGCATTTGTAAGTTGGCACTTGGAAATACAGTCTCAA 431
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
599 snIleValSerGlyThrAlaThrLeuGlyAlaAsnSerAlaThrThrAsp 615
432 GCAAGCTCACAACCTTGACCTGCTCAGCTCCGCTGCTGCCAATGC 481
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
616 AlaAsnGlyLysAla.....ThrValThrLeuLysSerSer..... 627
482 TACCCTGCCTCCCAATTACAGTACTTCCAAAACGACCAAGCACACCA 531
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
628 ThrProGlyGlnValValSerAlaLysThrAlaGluMetThrSerA 644
532 AATTCACCAAGCTCTGGTACTTATGCAATATTCGCCAAGGAGCTCC 581
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
644 lAlaLeuAsnAlaSerAlaValIlePheValGlnThrLysAlaSerIle 660
582 CCAATTCTCAGGGCGAGTGTCAAGCGCTCATTTGAATCACTGAAT...GG 628
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
661 ThrGluIleLysAlaAspLysThrThrAlaValAlaAsnGlyAsnAspAl 677

```

Tue Apr 2 09:39:55 2002

```
629 AAAACACAGTTACCTGGAGTACTGGATATATGGAGCAGGTGCTGATGCTA 678
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
677 aValThrTyrThrValLysValMetLysGluGlnProValHisGly. 693
679 CTAGGATGACGGTGTCTACTCAAGGTATTTCACAACCTTAT.....GAC 722
694 .....HisSerValAlaPheThrThrAsnPhcGlyMet 704
723 ACGAATGCTAGATACAGTGTAAAGTGGCGCTCTGGAGGAGTAAACGC 772
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
705 PheAsnGlyLysSerGlnThrGlnAsnAlaThrThrGlySerAspGlyAr 721
773 ACCCAGAGGAGATGATACCCAGCAGAGTGGAGCACTGTACATACCTG 822
721 gAlaThrIleThrLeuThrSerSerAlaGlyLysAlaThrValSerA 738
823 GCTGGATTGAGATGATGAATACAAATGGAATCCACCAAGACCTGAAAT 872
738 laThrValSerGlyAsn.....AspValLysAlaProGluVal 751
873 .....AATAAGGATGATGTTCAACACAA 895
752 ThrPhePheAspGlyLeuLysIleAspAsnLysValAspIleLeuGly 768
896 GCAAGTGTCT.....TTCA 909
768 sAsnValThrGlyAspLeuProAsnIleThrLeuGlnThrGlyGlnPhe 785
910 GCAGAACATCTCCGGAGGCTCATTTGTGGCTCT.....GATGTCACAA 956
785 ysLeuLysValSerGlyGlyAsnGlyThrTyrSerTriphHisSerGlu 801
957 GCTCCCATCTACCTGACTCTTCCACCTGCCAAATCACCGACCTCAAGC 1006
802 ThrAsnIleAlaThrValAspGluSerGlyLysValThr...LeuLys 817
1007 GGAATTCACGGGGGAGTCTCATTAATCTGACTGGACAGCTCTGGGG 1056
817 yLys.....GlyThrAlaValIleAsnValThr.....SerGlyA 829
1057 ATGATATTACCATGGA.....ACAGCTCACAGTATATCATTCGA 1097
829 sPlysGlnThrValSerTyrThrIleLysAlaProAsnTyrMetIleArg 845
1098 ATAAGTACAAGTATTCTTGATCTCAGACACAAGTTCATGAATCTCTTCA 1147
846 ValGly.....AsnLysAlaSerTyr 852
1148 AGTGAATACCTACTGCTCTCATCCCAAGGAGCCAACTCTGAGGAAGTCT 1197
852 rAlaAsnAlaMetSerPheCysGlyAsnLeuLeuProSerSerGlnThrV 869
1198 TTTTGTTTAAACCAAGAAACATTACTTTT...GAAATGCGACAGATCTT 1244
869 allLeuSerAsnValTyrAsnSerTyrGlyProAlaAsnGlyTyrAspHis 885
1245 TTCATTCTCTATTCAGGCTGTGTATAGGTGCTGATCTGAAATCAGAAATATC 1294
886 TyrArgSerMetGlnSerIleThrAlaTrpIleThrGlnThrGluAlaAs 902
1295 CAACATTCACGAGTACTTTGTTTATCTCTCCACAGACTCCGCCAGAGA 1344
902 pLysIleSerGlyValSerThrThrTyrAspLeuIleThr.....GlnA 917
1345 CACCTAGTCTCTGAGAACCTGCTCTCTGCTCTTAATATTCAT 1388
917 snProHisLysAsp...ValSerLeuAsnAlaProAsnValTyr 930
seq_name: sp_invertebrate:Q917U4
seq_documentation_block:
ID Q917U4 PRELIMINARY; PRT: 6815 AA.
```

```
AC Q917U4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG18242 PROTEIN.
GN CG18242.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyroidae; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Batllew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Besson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Duubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flosser C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kiamel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Peinado K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shae B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J., Yang Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Zhao Q., Zheng L.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
FL Science 287:2185-2195(2000).
CC -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC .. DOMAIN.
DR EMBL: AE003473; AAC22226.1; -.
DR FlyBase; FBgn0035301; CG18242.
DR InterPro; IPR003962; PnIII_repeat.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003600; Ig_Like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00047; Ig; 11.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SM0060; FN3; 2.
DR SMART; SM00408; IGC2; 5.
DR SMART; SM00410; IG_Like; 6.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
DR Repeat.
```

```
SQ SEQUENCE 6815 AA; 779559 MW; A4E2440014EBA01 CRC64;

alignment_scores:
  Quality: 122.00      Length: 554
  Ratio: 0.484        Gaps: 29
  Percent Similarity: 45.487  Percent Identity: 20.397

alignment_block:
US-09-049-696-19 x Q9I7U4 ..
Align seg 1/1 to: Q9I7U4 from: 1 to: 6815

39 CCCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAAATGACAGGAGTTT 88
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5271 ProGluAspAlaGlyGlnPheThrValValAlaLysAsnAlaGlyPhe 5287
89 ACAGACATATGCTTCAGATCAAGTTCAGAACAAATGGCTCATTTGATGCTT 138
||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
5287 eThrSerSerThrThrGluLeuValGluSerProLeuSerAspHis. 5303
139 TTGGGGCCCTTCATCAGGAAATGGAGCTGTCTCAGCGCTCCATCCAG 188
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5304 .....GlySerAspAlaThrAlaLeuSerArgArgSerMetSer 5316
189 CTTTCAGAGT.....AAGGATTAAACCTCCAGAACAG 220
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5317 ArgGluSerSerLeuAlaAspIleLeuGluGlyIleProProThrPheSe 5333
221 CCAGTGGATGAATGGCACAGTGTGGACAGCAGCGTGGGAAAGGACA 270
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5333 rLysLysProLysAlaGlnTyrValAspGluAsnThr.....AsnV 5347
271 CTTTGTTCCTATACCTGAGCAACGAGCGCTCCCAAAATCCTCTCTGG 320
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5347 alIleLeuGluCysArgLeuValAlaValProGluProAspIleValTrp 5363
321 GATCCCGAGTGGACAG.....AAGCAAGGTGGCTT 349
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5364 ThrPheAsnGlyGluAspIleAspGluGluGluIleLysAsnValArgII 5380
350 TGTAGTGGCAAAACACCAAAATGGCTACCTCCAAATCCCGAGGATTTG 399
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5380 eValThrGluSerAspMetHisMet...TyrCysSerValValHisIleS 5396
400 CTAAGGTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAAACCTTG 449
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5396 erLysVal.....LysLysSerGlnGluGlyThrTyr 5406
450 ACCCTGACTGTACGTCGCGTGGCTCAATGCTACCTGCTCCCAATTAC 499
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5407 GluValIleAlaThrAsnArgGluGlyGluAlaArgLeuProIleThrLe 5423
500 AGTGACTTCCAAACGACCAAGCACACCAGCAAAATCCCGAGCGCTCTGG 549
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5423 uLysValArgThrThrAspLysGluAlaProGlnIleLeuGluProLeu. 5439
550 TAGTTTATGCAAT.....ATTCCGCAAGGACCGCTCCCAATTCACAGG 593
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5440 .....ArgAsnMetValIleArgGluGlyGluSerValValLeuSer 5453
594 GCCAGTGTACAGCC.....CTGATTGAATCAGTGAATGG 628
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5454 ThrGlnIleValGlyAsnProProProLysValThrTrpTyrLysAspGI 5470
629 AAAACAGATTACCTTGGAACTACTGGATAATGAGCAGGCTGCTCATGCTA 678
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5470 yLysProVal.....LysAsnAlaLysSera 5479
679 CTAAGGATGACGGTGTCTACTCAAGGTATTTCAACATTATGACAGAAAT 728
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5479 sPLysAspLeuHisThrLeuThrLeuIleThrProGlnLysSerGluLys 5495
```

```
723 GGTAGATACAGTGTAAAA...GTGCGGGCTCTGGAGGAGTAAACCGCAGC 775
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5496 GlyGluTyrThrValLysAlaValAsnProLeuGlySerValGluThrTh 5512
726 CAGACGGAGAGTATACCCAG..... 797
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5512 rAlaAsnLeuThrIleGluProAlaGlyGlyAsnAlaGluProProL 5529
728 ..... 798
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5529 euPheValGluArgPheGluGluGlnAsnValProGlnLysGlyGluIle 5545
813 TACATACCTGGCTGGATTGACAAATGAT.....GAAATCAATAGG... 851
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5546 ArgLeuProAlaLysValSerGlyAsnProValProGluValGlnTrpLe 5562
852 .....AATCCACCA..... 860
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5562 uPheAsnAsnThrProLeuPheProSerGluArgIleGlnGlnValTyrA 5579
861 .....AGACCTGAAATTAAT 875
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5579 spGlyGluAsnIleGluLeuIleIleLysAspAlaAsnProGluThrAsp 5595
876 AAGGATCATCTTCAACACAGCAAGTGTGTTTCAGCAGAACA..... 917
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5596 SerGlyAspTyrLys.....CysIleAlaSerAsnProIleGI 5608
918 ....TCTCGGGAGAGCTCATTTGTGCTTCTGATGTGCCCAATGCT.... 959
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5608 yLysThrSerHisGlyAlaArgValIleValGluValAspGluValThrP 5625
960 .....CCATACCTGATCTCTTCCACCTGGCCAAATCACCACCTCC 1001
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5625 heThrLysLysLeuLysLysThrIleThrIleGluGluValGlnSerLeu 5641
1002 AAGCGCGAAATTCACGGGGCAGCTCTCAATTAATCTGACTTGGACAGCTCC 1051
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5642 ThrLeuGluCysGluThrSerHisValValThrThrLysTrpPhePheAs 5658
1052 TGGGGATGAT.....TATGACCAT.....GGAA 1074
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5658 nGlyLysGluLeuSerGlyMetAspHisArgValValValGluAspGlyL 5675
1075 CAGCTCACAGTATATCATTCGAATAGTACAAGTATCTTGTATCTCAGA 1124
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5675 yThrHisLysLeuValIleArgAsnThr.....AsnLeuArg 5687
1125 GAC..... 1127
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5688 AspSerGlyThrTyrThrCysLysValLysLysGlnGluThrGlnSerTh 5704
1128 .....AAGTTCAATCAATCTC 1143
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5704 rValGluValLeuGlnArgLysProAspPheIleLysValLeuGluAspT 5721
1144 TTCAGTGT.....AATACTACTGCTCTCATCCCAAGGAAGCAACTCT 1187
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5721 yGluValThrGluLysAspThrAlaIleLeuAspValGluLeuThrThr 5737
1188 GAG.....GAAGTCTTTTGTTTAAACCA...GAAACATTACTTTTGA 1238
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5738 GluAlaThrGluValThrTrpTyrLysAspGlyLysIleThrProGI 5754
1229 AATGCCACAGATCTTTTCATT.....GCTATTACGCGCTGTTG 1266
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5754 uAsnLysAsnValGluPheIleLysAspGlyLysAla.ArgArgLeuVal 5770
1267 ATAAGTCTCATCTGAAATCAGAAATATCCAACTTGCAC.....GA 1307
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5771 IleArgAspVal.....ThrIleHisAspGluGlyGI 5781
1308 GTATCTTTGT 1317
```

271 CTTTGTTCCTTATCACCTGGCAACACGAGCCTCCCAATCCCTCTCTCGG 320
14716 alIleLeuGluCysArgLeuValalaValProGluProAspIleValTrp 14732
321 GATCCCACTGGACAG.....AAGCAAGGTGGCTT 349
14733 ThrPheAsnGlyGluAspIleAspGluGluGluIleLysAsnValArgI 14749
350 TGTAGTGCACAAAAACCAAAATGGCTACCTCCCAAAATCCCAAGCATTG 399
14749 eValThrGluSerAspMetHisMet...TyrCysSerValValHisIle 14765
400 CTAAGGTTGGCACTTGGAAATACAGTCTGCAAGCAGCTCACAAACCTTG 449
14765 erLysVal.....LysLysSerGlnGluGlyThr 14775
450 ACCCTGACTGTACGTCGCTCGCTGCCTCAAGTCTACCTGCCTCCAATTAC 499
14776 GluValIleAlaThrAsnArgGluGlyGluAlaArgLeuProIleThrIe 14792
500 AGTGACTTCCAAAACAAACAGGACACACGCAAAATCCCAAGCCCTCG 549
14792 uLysValArgThrAspLysGluAlaProGlnIleGluProLeu. 14808
550 TAGTTTATGCAAAAT.....ATTGCCAAGAGAGCCTCCCAATTCACAG 593
14809ArgAsnMetValIleArgGluGlySerValValLeuSer 14822
594 GCCAGTGTACAGCC.....CTGATTCAATCACTGAATGG 628
14825 ThrGlnIleValGlyAsnProProLysValThrTrpLysAspGI 14839
629 AAAAACACTTACCTTGGAACTACTGGATAATPAGGACAGGTGCTGATGCTA 678
14839 yLysProVal.....LysAsnAlaLysSerA 14848
679 CTAAGGATCAGCGTCTCTACTCAAGGTATTTCAACAATTCATGACACCAAT 728
14843 sPlAspLeuHisThrLeuThrLeuIleThrProGlnLysSerGluLys 14864
729 GGTAGATACAGTGTAAAA...GTCCGGCTCTGGAGAGGTAAACGAGC 775
14865 GlyGluTyrThrValLysAlaValAsnProLeuGlySerValGluThrTh 14881
776 CAGACGGAGAGTGATACCCAG..... 797
14881 rAlaAsnLeuThrIleGluGluProAlaGlyClyAsnAlaGluProProL 14898
793CAGAGTGGAGCACTG 812
14898 euPheValGluArgPheGluGluGlnAsnValProGlnLysGlyGluIle 14914
813 TACATACCTCGCTGGATTCAGAAATGAT.....GAATACAAATCG... 851
14915 ArgLeuProAlaLysValSerGlyAsnProValProGluValGlnTrpLe 14931
852AATCCACCA..... 860
14931 upheAsnAsnThrProLeuPheProSerGluArgIleGlnInValTyrA 14948
861AGACCTGAATTAAT 875
14948 sPGlyGluAsnIleGluLeuIleLysAspAlaAsnProGluThrAsp 14964
875 AAGATGATGTTCAACACAGCAAGTGTCTTTCACGACAACA..... 917
14965 SerGlyAspTyrLys.....CysIleAlaSerAsnProIleGI 14977
918TCTCTGGAGGCTCATTTGTGCTTCTGATGTCCTCAATGCT.... 959
14977 yLysThrSerHisGlyAlaArgValIleValGluValAspGluValThr 14994

```

776   CAGACGGAGGTGATACCACCG...    797
      :       :       :       :
14861 rAlaAsnLeuThrIleGluGluProAlaGlyCysAlaGluProProL 14898
      :       :       :       :
793   .....CAGCTGGAGCACTG 812
      :       :       :       :
14898 euPheValGluArgPheGluGluGlnAsnValProGlnGlyGlyLeu 14914
      :       :       :       :
      :       :       :       :
813   TACATACCTCGGCTGGATTGAGAATCAT.....GAATACACAATGC... 851
      :       :       :       :
14915 ArgLeuProAlaLysValSerGlyAsnProValProGlnGlyGlnTrpLe 14931
      :       :       :       :
      :       :       :       :
852   ....AATCCACCA.....    860
      :       :       :       :
14933 uPheAsnAsnThrProLeuPheProSerGluArgIleGlnGlnValTyra 14948
      :       :       :       :
      :       :       :       :
861   .....AGACTTCGAATTAAT 875
      :       :       :       :
14948 spGlyGluAsnIleGluLeuIleLysAspAlaAsnProGluThrAsp 14964
      :       :       :       :
      :       :       :       :
875   AAGATGATGTTCAACACAAAGTGCTGTTTCCACGAACA.....    917
      :       :       :       :
14965 SerGlyAspTyrLys.....CysIleAlaSerAsnProIleGL 14977
      :       :       :       :
      :       :       :       :
918   ...TCCTCGGAGGCTCATTTGTGGCTTCTGATGTCCTCCAATGCT.... 959
      :       :       :       :
      :       :       :       :
14977 vMembrSerHisGlnAlaArvAlIleValGluValAspGluValThr 14994
```

```
960 .....CCCATACGTGATCTCTCCACCTGGCCAAATCAACCGACCTG 1001
14994 heThrLysLysLeuLysLysThrIleGluGluValGlnSerLeu 15010
1002 AAGCGGAAATTCACGGGGCAGCTCTCATTAATCTGACTTGGACAGCTCC 1051
15011 ThrLeuGluCysGluThrSerHisValValThrThrLysTrpPhePheas 15027
1052 TGGGGATGAT.....TATGACCAT.....GGAA 1074
15027 nGlyLysGluLeuSerGlyMetAspHisArgValValGluAspGlyL 15044
1075 CAGCTCACAGTATATCATTCGATACAGTACAGTATCTTGTGATCTCAGA 1124
15044 ystThrHisLysLeuValLeuArgAsnThr.....AsnLeuArg 15056
1125 GAC..... 1127
15057 AspSerGlyThrThrCysLysValLysLysGlnGluThrGlnSerTh 15073
1128 .....AAGTTCAATGAATCTC 1143
15073 rValGluValLeuGlnArgLysProAspPheIleLysValLeuGluAspT 15090
1144 TTCAAGTG.....AATACTACTGCTCTCATCCCAAGGAGCAACTCT 1187
15090 yrgLuuValThrGluLysAspThrAlaIleLeuAspValGluLeuThrThr 15106
1188 GAG.....GAAGCTCTTTTGTGTTAAACCA...GAAACATTACTTTTGA 1228
15107 GluAlaThrGluValThrTrpTyrLysAspGlyGluLysIleThrProG 15123
1229 AATGGCAGAGATCTTTTCATT.....GCTATTCAGCGTGTTC 1266
15123 uAsnLysAsnValGluPheIleLysAspGlyLysAla.ArgArgLeuVal 15139
1267 ATAAGGTGCATCGAATCAGAAATATCCCAACATTCAC 1307
15140 IleArgAspVal.....ThrIleHisAspGluGlyG 15150
1308 GTATCTTTGT 1317
15150 nTyrThrCys 15153
```

seq_name: sp_invertebrate:Q9UB29

```
seq_documentation_block:
AC Q9UB29; PRELIMINARY; PRT; 4280 AA.
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DE MYOACTIN FORM A (H19M22.2B PROTEIN).
GN LET-805 OR H19M22.2B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RA Hresko M.C., Schriefer L.A., Shrimankar P., Waterston R.H.;
RT "Myotactin, a novel cell-adhesion protein associated with
RT hemidesmosome-like structures in Caenorhabditis elegans.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
```

```
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Wilson R., Wamsley P.;
RT "The sequence of C. elegans cosmid H19M22.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF148954; AAD37411.1;
DR EMBL: AF040648; AAK21414.2;
DR HSSP: P24821; ITEN.
DR IngerPro; IPR003962; FNIII_repeat.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00041; fn3; 32.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SM00060; FN3; 24.
KW Repeat.
SQ SEQUENCE 4280 AA; 479552 MW; 9BA57F89C328090E CRC64;
alignment_scores:
Quality: 121.50 Length: 542
Ratio: 0.547 Gaps: 30
Percent Similarity: 40.959 Percent Identity: 19.926
alignment_block:
US-09-049-696-19 x Q9UB29
Align seg 1/1 to: Q9UB29 from: 1 to: 4280
195 AGTAAGGATTAACTCCAGAACAGCCAGTGGATGAATGGCACAGTGAT 244
RL SerLysGluLeuSerIleThrThrSerGluGlyAlaProGlyPro..... 2524
245 CGTGGACAGCACCGTGGGAAG.....GACACTTTGTTTCTTATCA 285
2525 ValGluGlyLeuThrAlaLysProThrGlyProThrSerIleValAla 2541
286 CCGTGACAGCAGCTCCCAAACTCTCTCTGGATCCAGTGGACAG 335
2541 rGrp.....LysProProArg.....AspProAsnGlyVal 2551
336 AGCAAGGT..... 344
2552 IleThrGlyTyrThrLeuThrTyrGlnLeuLysSerIleGlyGluCysG 2568
345 .....GGCTTTGTAGTGACACCAACACCAAAATGGCTACCTCC 384
2568 yProArgSerSerThrProIleGluLysHisValLysAsnGluGluG 2585
385 AAATCCAGCAGCTTGTAAAGTTGGCATTGGAAA..... 419
2585 hrLeuGluGlyLeuLeuProAspSerThrTyrGluIleHisValAla 2601
420 TACAGTCTCAAGCAAGCTCACAAACCTTGACCTGACTGTCACGTCCCG 469
2602 HistThrSerHisAlaGlyProGlnSerSerValValThrValThrG 2618
470 TGGCTCCATGCTACCTGCCTCCAATATTACAGTACTTCCAAACGAACA 519
2618 uGluAlaAlaProThrGlyProPro..... 2626
520 AGCACACCAGCAATATCCCGACCCCTCTGCTAGTTATGCAATATTCG 569
2627 .....GlnAsnValArg 2630
570 CAAGGAGCCTCCCAATATTCAGGGCCAGTGTCACA.....GCCCT 610
2631 AlaGlySerValThrSerSerArgAlaAspIleThrTrpThrGlnProG 2647
```

assignment_scores:	Quality:	121.50	Length:	542
--------------------	----------	--------	---------	-----

Length: 542


```
seq_documentation_block:
ID Q10466 PRELIMINARY; PRT; 26926 AA.
AC Q10466;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE TITIN, HEART ISOFORM N2-B (EC 2.7.1.-) (CONNECTIN).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RX MEDLINE=96026330; PubMed=7569978;
RA Labelit S., Kolmer B.;
RT "Titins: giant proteins in charge of muscle ultrastructure and
RT elasticity.";
RL Science 270:293-296(1995).
RN [2]
RP SEQUENCE OF 22277-25376 FROM N.A.
RX MEDLINE=92258380; PubMed=1582406;
RA Labelit S., Gautel M., Lakay A., Trinick J.;
RT "Towards a molecular understanding of titin.";
RL EMBO J. 11:1711-1716(1992).
RN [3]
RP SEQUENCE OF 1976-2014 FROM N.A.
RA Labelit S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION
RX MEDLINE=95331314; PubMed=7607248;
RA Gautel M., Castiglione-Morelli M.A., Pfuhl M., Motta A., Pastore A.;
RT "A calmodulin-binding sequence in the C-terminus of human cardiac
RT titin kinase.";
RL Eur. J. Biochem. 230:752-759(1995).
CC -!- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE
CC ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF
CC SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.
CC -!- ALTERNATIVE PRODUCTS: A NUMBER OF FORMS OF THIS PROTEIN ARE
CC PRODUCED BY ALTERNATIVE SPLICING WHICH DIFFER IN TISSUE
CC DISTRIBUTION. DIFFERENT SIZE TRANSCRIPTS MAY ALSO EXIST WITHIN ANY
CC ONE TISSUE. THE SEQUENCE SHOWN HERE IS THAT OF THE HEART ISOFORM
CC N2-B.
CC -!- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
CC -!- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE
CC KINASES.
CC -!- SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 112
CC IMMUNOGLOBULIN C2-LIKE DOMAINS AND 132 FIBRONECTIN TYPE III-LIKE
CC DOMAINS.
CC EMBL; X64698; CAA45939.1; -
CC EMBL; X83270; CAA58243.1; -
CC EMBL; X64697; CAA45938.1; -
CC EMBL; X90568; CAA62188.1; -
CC EMBL; X64699; CAA45940.1; -
CC HSP; P56276; ITLK.
CC InterPro; IPR000282; Cytok_receptor_2.
CC InterPro; IPR000719; Euk_kinase.
CC InterPro; IPR000577; FG_Y_kin.
CC InterPro; IPR003962; FhIII_repeat.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR003015; HLH_Myc.
CC InterPro; IPR003598; Ig_c2.
CC InterPro; IPR003600; Ig_like.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR000129; Peptidase_S24.
CC InterPro; IPR002016; Peroxidase.
CC InterPro; IPR002290; Ser_thr_kin_actsite.
CC InterPro; IPR001245; Tyr_kin.
CC Pfam; PF00041; fn3; 132.
CC Pfam; PF00047; Ig; 95.
CC Pfam; PF00069; pkinase; 1.
CC PRINTS; PR00014; FNTPYELII.
```

```
DR PRINTS; PR00726; LEXASERPTASE.
DR SMART; SM00060; FN3; 127.
DR SMART; SM00408; IGC2; 23.
DR SMART; SM00410; IG_Like; 79.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00933; FG_Y_KINASES.1; UNKNOWN1.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN1.
DR PROSITE; PS00290; IG_MHC; UNKNOWN1.
DR PROSITE; PS00435; PEROXIDASE.1; UNKNOWN1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Muscle protein; Cytoskeleton; Structural protein; Calmodulin-binding;
KW Serine/threonine-protein kinase; Alternative splicing; Repeat;
KW Immunoglobulin domain; Phosphorylation.
KW DOMAIN 1370 1389 4 X 5 AA TANDEM REPEATS OF R-M-S-P-A.
FT DOMAIN 4429 4614 GLU/LYS/PRO/VAL-RICH.
FT DOMAIN 24731 25070 CATALYTIC.
FT DOMAIN 25030 25056 CALMODULIN-BINDING.
FT MOD_RES 1372 1372 PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT MOD_RES 1377 1377 PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT MOD_RES 1382 1382 PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT MOD_RES 1387 1387 PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT MOD_RES 26171 26171 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 26178 26178 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 26184 26184 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 26190 26190 PHOSPHORYLATION (POTENTIAL).
FT CONFLICT 22277 22277 T -> P (IN REF. 2).
FT CONFLICT 22449 22449 E -> G (IN REF. 2).
FT CONFLICT 22454 22454 T -> Q (IN REF. 2).
FT CONFLICT 23324 23324 S -> L (IN REF. 2).
SQ SEQUENCE 26926 AA; 2993428 MW; D5EEDC3254DF5523 CRC64;
```

```
alignment_scores:
Quality: 121.00 Length: 517
Ratio: 0.545 Gaps: 29
Percent Similarity: 42.940 Percent Identity: 19.729

alignment_block:
US-09-049-696-19 x Q10466
Align seg 1/1 to: Q10466 from: 1 to: 26926

145 CCCTTTCATCAGGAATGGAGCTGCTCTCAGCGCTCCATCCAGCTTGAG 194
||| : : : : : : : : : : : : : : : : : : : : : : : : :
7073 ProAlaTyrValAspGluProValAsnMetSerThrProAlaThr..... 7089

195 AGTAAGGGATTAACTCCAGAACACCCAGTCAGTGAATGGACAGTGAT 244
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7090 .ValProAspProPro..GluAsnValLysTrpArgAspArgThrAla.. 7104

245 CGTGGACAGCACCGTGGGAAAGGACACACTTTGTTTCTATCACCCTGGACAA 294
: : : : : : : : : : : : : : : : : : : : : : : : :
7105 .....AsnSerIlePheLeu...ThrTrp.... 7111

295 CGCAGCTCCCAAAATCTTCTCTGGGATCCACAGTCAGACAGAAAGGT 344
: : : : : : : : : : : : : : : : : : : : : : : : :
7112 ..AspProProLys.....AsnAspGlyGlySerArgIleLys 7123

345 GCGTTTGTAGTGGACAAA..... 362
||| : : : : : : : : : : : : : : : : : : : : : : : : :
7124 GlyTyrIleValGluArgCysProArgGlySerAspLysTrpValAlaCy 7140

363 .....AACACCAAAATGGCTTACCTCCCAAAATCCCAAGGCA 396
: : : : : : : : : : : : : : : : : : : : : : : : :
7140 sGlyGluProValAlaGluThrLysMet.....GluValThrGlyL 7154

397 TTGCTAAGTTGGCACTTGG.....AAATACACTGTG 428
: : : : : ||| : : : : : : : : : : : : : : : : : :
7154 euGluGlu...GlyLysTrpTyrAlaTyrArgValLysThrLeuAsnArg 7169

429 CAAGCAAGCTCACAA..... 443
||| : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
7170 GlnGlyAlaSerLysProSerArgProThrGluGluIleGlnAlaValAs 7186
444 .....ACCTTGACCCCTGACTG 459
7186 pThrGlnGluAlaProGluIlePheLeuAspValLysLeuAlaGlyL 7203
460 TCAGCTCCCGTCCGCTCCAAATGCTACCTGCTCCAAATACAGTACCTCC 509
7203 euThrValLysAlaGlyThrLysIleGluLeuProAlaThrValThrGly 7219
510 AAACAGCAAGGACACACAGCAAAATCCCGAGCCCTCTGGTAGTTATGC 559
7220 Lys.....ProGluProLysIleThrIleThrIle 7228
560 AAATATTGCCAAGGAGCTCCCAATCTCAGGGCCAGTGTACAGCC 609
7228 rLys.....AlaAspMetIleLeuLysGlnAspLysArgIleT 7241
610 TGATTGAATCAGTGAATGGAAAA...ACAGTTACCTTGGAACTACTGGAT 656
7241 hrIleGluAsnValProLysLysSerThrValThrIle..... 7253
657 ANTGGAGCAGTGTGCTGCTACTAAGGATGACGGTGTCTACTCAAGTA 706
7254 .....ValAspSerLysArgSerAsp..... 7260
707 TTTCACACTTATGACACGAATGGTAGATACAGTGTAAAGTGGGGCTC 756
7261 .....ThrGlyThrTyrIleIleGlu..... 7267
757 TGGGAGGAGTTAAACGACCCAGAGGAGGATGATACCCACGAGAGTGA 806
7268 .....AlaValAsnValCysGlyArgAlaThrAlaValAlaGluValAsn 7282
807 GCACGTGTACATACCTGGC.....TGGATTGA 832
7283 ValLeuAspLysProGlyProProAlaAlaPheAspIleThrAspValTh 7299
833 GAATGATGAA.....ATACATGAATCCACACAGACCTGAA..... 869
7299 rAsnGluSerCysLeuLeuThrTrpAsnProProArgAspAspGlyGly 7316
870 .....ATTAATAAGGATGATGTT 887
7316 exLysIleThrAsnTyrValValGluArgAlaThrAspSerGluVal 7332
888 CAACACAGCAAGTGTGTTTCAGCAGAACATCCTCGGGAGGCTCATTTGT 937
7333 TrpHisLys.....LeuSerSerThrValLysAspThrAsnPhely 7346
938 GGCCTTCTGAT...GTCCCAAT..... 956
7346 salathrLysLeuIleProAsnLysGluTyrIlePheArgValAlaAlaG 7363
957 .....GCTCCCATACCTGAT 971
7363 luAsnMetTyrGlyAlaGlyGluProValGlnAlaSerProIleThrAla 7379
972 CTCCTTC.....CCACCTGGCCAAATCACCAGCTGAGGGCGGAAT 1012
7380 LysTyrGlnPheAspProGlyProProThrArgLeuGluProSerAs 7396
1013 TCACGGGGCAGCTCTATTGATTCAGTGGACAGCTCCTCGGGGATGATT 1062
7396 pIleThrLysAspAlaValThrLeuThrTrpCysGluProAspAspAspG 7413
1063 ATGACCATGGAACAGCTCACAAGTATATCATTTCAATAAGTACAAGTATT 1112
7413 LysGlySerProIleThrGlyTyrTrpValGluArgLeuAspProAsp... 7428
1113 CTTGATCTCAGACACAAAGTTC.....ATGAATCTCTCAAGTGNATAC 1156
7429 .....ThrAspLysTrpValArgCysAsnLysMetProValLysAs 7442
```

```
1157 TACTGCTCTCATCCCAAGGAAGCAACTCTCAGGAAGTCTTTTGTGTTTA 1206
7442 pThrThrTyrArgValLysGlyLeuThrAsnLysLysLysTyrArgPheA 7459
1207 AACCAAGAAACATTACTTTTGAATGGCACAGATCTTTTCATTGCTATT 1256
7459 rg.....ValLeuAlaGluAsnLeuAlaGlyProGlyLysProSer 7472
1257 CAGGCTGTTGATAAGTCTGATCTGAAATCAGAAATATCCAAACATTGCAG 1306
7473 LysSerThrGluProIleLeuIleLysAspProIleAsp..... 7485
1307 AGTATCTTTGTTATTCTCCACAGACTCCGCCAGACACACCTAGTCTCTG 1356
7486 .....ProProTrpProProGlyLysProThrValL 7496
1357 ATGAAACGTCCTGCTCTGCTCTTAATATTCATATCAACAGCACCATTCCT 1406
7496 ysaspValGlyLys.....ThrSerValArgLeuAsnTrpThrLysPro 7510
```

OM of: US-09-049-696-19 to: SwissProt_39.* out_format : pfs
 Date: Mar 30, 2002 2:56 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=framed_n2p.model -DEV=xlpl
 -Q/cgn2_1/USPTO.spool/US09049696/runat_28032002_145238_2085/app_query.fasta_1.12579
 -DB=SwissProt_39 -QFMT=fastan -SUFFIX=isp -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -DELOP=6.000 -FEAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM-ext -MINLEN=0 -MAXLEN=200000000
 -USER=US09049696 @CGN1_1_165 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
 Query: US-09-049-696-19
 Query length: 1683
 Database: SwissProt_39.*
 Database sequences: 100059
 Search time (sec): 306.030000

score_list:	Strd Orig	ZScore	Escore	Len	Documentation
SwissProt_39:PGMB_HUMAN	132.00	144.07	0.0539	4393	! P98160 homo sapiens (human)
SwissProt_39:YPR3_CAEEL	131.00	160.24	0.0434	686	! Q20059 caenorhabditis elegans
SwissProt_39:PMR2_CHLUM	130.00	150.53	0.0618	1672	! O9PIY2 chlamydia muridarum
SwissProt_39:NEO1_MOUSE	124.50	144.36	0.1529	1493	! P97798 mus musculus (mouse)
SwissProt_39:EAEL_ECOLI	124.00	148.11	0.1510	935	! O31000 escherichia coli o111:
SwissProt_39:PTPD_HUMAN	124.00	141.37	0.1752	1912	! P23468 homo sapiens (human)
SwissProt_39:MAG1_YEAST	120.50	151.10	0.2306	417	! P35192 saccharomyces cerevis
SwissProt_39:SPG7_DICDI	119.50	147.49	0.2872	532	! P22698 dictyostelium discoid
SwissProt_39:IGM4_HAEN	117.50	133.13	0.5214	1849	! P45386 haemophilus influen
SwissProt_39:DCC_HUMAN	115.00	132.14	0.7559	1447	! P43146 homo sapiens (human)
SwissProt_39:GUXA_CELFI	114.50	136.25	0.7404	872	! P50401 cellulosomonas fimi. ex
SwissProt_39:MPC2_MYCPN	114.00	132.45	0.8635	1218	! Q50341 mycoplasma pneumonia
SwissProt_39:NEO1_RAT	113.50	130.64	0.9638	1377	! P97603 rattus norvegicus (rat
SwissProt_39:EAEL_ECO57	111.50	131.66	1.125	934	! P43261 escherichia coli o157:
SwissProt_39:NRG_DROME	110.50	127.21	1.58	1302	! P20241 drosophila melanogast
SwissProt_39:PMPC_CHLUM	110.50	126.13	1.62	1460	! O9PIY1 chlamydia muridarum
SwissProt_39:FLIC_ECOLI	110.00	135.62	1.41	497	! P04549 escherichia coli. fla
SwissProt_39:YRM8_CAEEL	110.00	122.79	1.87	1941	! Q09417 caenorhabditis eleg
SwissProt_39:DCC_MOUSE	107.50	122.27	2.68	1447	! P70211 mus musculus (mouse)
SwissProt_39:EP33_BRARE	107.00	125.27	2.69	981	! O13146 brachydanio rerio (ze
SwissProt_39:GUXB_CELFI	107.00	124.28	2.73	1090	! P50899 cellulosomonas fimi. e
SwissProt_39:FINC_CHICK	107.00	122.94	2.83	1256	! P11722 gallus gallus (chick
SwissProt_39:PTPE_HUMAN	107.00	119.06	3.09	1897	! P10586 homo sapiens (human)
SwissProt_39:NEO1_HUMAN	106.50	120.86	3.18	1461	! Q92859 homo sapiens (human)
SwissProt_39:NRP_STRSU	106.00	121.63	3.36	1256	! P32653 streptococcus suis. m
SwissProt_39:NCO2_RAT	106.00	120.18	3.46	1465	! Q9WU19 rattus norvegicus (rat
SwissProt_39:PGCA_CHICK	106.00	116.75	3.74	2109	! P07898 gallus gallus (chick
SwissProt_39:POXA_YEEN	105.50	126.34	3.24	710	! Q01674 versinia enterocoliti
SwissProt_39:IGA2_HAEN	105.50	118.11	3.89	1702	! P45384 haemophilus influen
SwissProt_39:IL6B_RAT	105.00	123.26	3.72	918	! P40190 rattus norvegicus (rat)
SwissProt_39:IL6B_MOUSE	104.50	122.61	4.05	917	! Q00360 mus musculus (mouse)
SwissProt_39:MFD_MYXXA	104.50	120.07	4.28	1201	! O52236 myxococcus xanthus. t
SwissProt_39:P302_HUMAN	104.00	117.55	4.86	1464	! P15596 homo sapiens (human)
SwissProt_39:PK31_DICDI	104.00	116.89	4.93	1570	! P54673 dictyostelium discoid
SwissProt_39:FINC_MOUSE	104.00	112.60	5.42	2477	! P11276 mus musculus (mouse)
SwissProt_39:DNM_MYCPN	103.00	120.71	5.21	910	! P73564 mycoplasma pneumoniae
SwissProt_39:NUAM_BOVIN	102.50	122.17	5.41	727	! P15690 bos taurus (bovine)
SwissProt_39:DSC4_HUMAN	102.50	112.58	6.68	2012	! O60469 homo sapiens (human)
SwissProt_39:RAD3_SCHPO	102.50	110.97	6.92	2386	! Q02099 schizosaccharomyces
SwissProt_39:LYTD_BACSU	102.00	119.71	6.12	880	! P39848 bacillus subtilis. be

SwissProt_39:CAL7_HUMAN + 102.00 108.34 7.87 2944 ! Q02388 homo sapiens (hum
 SwissProt_39:HOC_BPT4 + 101.50 127.06 5.59 376 ! P18056 bacteriophage t4. hi
 SwissProt_39:CAD6_CHICK + 101.50 120.07 6.52 790 ! O90762 gallus gallus (chi
 SwissProt_39:CD22_MOUSE + 101.50 119.24 6.64 862 ! P35329 mus musculus (mous
 seq_name: SwissProt_39:PCBM_HUMAN
 seq_documentation_block:
 ID PCBM_HUMAN STANDARD; PRT: 4393 AA.
 AC P98160; Q16287;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE
 DE PROTEIN PRECURSOR (HSPG) (PERLECAN) (PLC).
 GN HSPG2.
 OS Homo sapiens. (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RP [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=92112994; PubMed=1730768;
 RA Kallunki P., Tryggvason K.;
 RT "Human basement membrane heparan sulfate proteoglycan core protein: a
 RT 467-kD protein containing multiple domains resembling elements of the
 RT low density lipoprotein receptor, laminin, neural cell adhesion
 RT molecules, and epidermal growth factor.";
 RL J. Cell Biol. 116:559-571(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92112994; PubMed=1569102;
 RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
 RT "Primary structure of the human heparan sulfate proteoglycan from
 RT basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
 RT domains homologous to the low density lipoprotein receptor, laminin,
 RT neural cell adhesion molecules, and epidermal growth factor.";
 RL J. Biol. Chem. 267:8544-8557(1992).
 RN [3]
 RP SEQUENCE OF 1018-1472 FROM N.A.
 RX MEDLINE=91365376; PubMed=1679749;
 RA DeGee G.R., Kovalsky I., Chu M.L., Hassell J.R., McBride O.W.,
 RA Yi.H.F., Iozzo R.V.;
 RT "Heparan sulfate proteoglycan of human colon: partial molecular
 RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
 RT short arm of human chromosome 1.";
 RL Genomics 10:673-680(1991).
 RN [4]
 RP SEQUENCE OF 892-1398 FROM N.A.
 RX MEDLINE=92120660; PubMed=1685141;
 RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
 RA Tryggvason K.;
 RT "Cloning of human heparan sulfate proteoglycan core protein,
 RT assignment of the gene (HSPG2) to lp36.1-->p35 and identification of
 RT a BamHI restriction fragment length polymorphism.";
 RL Genomics 11:389-396(1991).
 RN [5]
 RP SEQUENCE OF 1-21 FROM N.A.
 RX MEDLINE=94052171; PubMed=8234307;
 RA Cohen I.R., Graessell S., Murdoch A.D., Iozzo R.V.;
 RT "Structural characterization of the complete human perlecan gene and
 RT its promoter.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
 CC -!- FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT
 CC MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
 CC CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
 CC PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
 CC SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
 CC SUBSTRATE FOR CELLS.
 CC -!- SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN


```
17 CATCCACACAGTCGCTTTGGGGCCCTCTCCAGCTCAAGAAGCTAGAGGACC 66
1621 AsnProGluAsnMetPheSerArgThrCysGluSerLeuGlyAlaGlyI 1637
67 TGTCCAAAATGACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCA 116
1637 yTyraG.....CysThrAlaCysGluProGlyTyThrG 1649
117 ACAATGGCCTCATGATGCTTTTGGGGC.....CCTTTCATC 154
1649 yGlnTyrcysGluGlnCysGlyProGlyTyrcysGlyValGlyAsnProSerVal 1665
155 AGGAATGGAGCTGCTCTCA.....GC 177
1666 GlnGlyGlyGlnCysLeuProGluThrAsnGlnAlaProLeuValValG 1682
178 GTCCATCAGCTTGAGAGTAAGGATTAAACCTCCAGAACAGCCAGTGG 227
1682 uValHisProAla.....ArgSerIleValProGlnG 1693
228 ATGAATGGCAGAGTATCGTGGACAGCACCCTGGGAAGGACACTTTGTT 277
1693 yGlySerHisSerLeuArgCysGlnValSerGlyArgGlyProHisTy 1709
278 TCTTATCAGCTGGACACAGCAGCCTCCCAAAATCTCTCTGGATCCCA 327
1710 PheTyrTrpSerArgGluAspGlyArgProValProSer...GlyThrG 1725
328 GTGGCAGAGAAGCAAGTGGCTTTGTAGTGGACAAAACACCAAAATGGCC 377
1725 n.....GlnArgHisGlnGly.Ser 1731
378 TACCTCCAAATCCAGGCAATTCCTAAGTTGGCAGCTTGGAAATACAGTCT 427
1732 GluLeuHisPheProSerValGlnProSerAspAlaGlyValTyIle.. 1747
428 GCAAGCAAGCTCACAAACCTTCACCTGACTGTCACGTCCTCCGTCG.. 473
1748 ....CysThrCysArgAsnLeuHisArgSerAsnThrSerArgAlaGlu 1763
474 .....TCCAATGCTACCCCTCCCAATTACAGTACTGACTTCCAAAAGC 515
1763 euLeuValThrGluAlaProSerLysProIleThrValThrValGluGlu 1779
516 AACAGGACACACAGAAATCCCGAGCCCTCTGGTAGTTTATGCAAAAT 565
1780 GlnArgSerGlnSerValArgProGlyAlaAspValThrPheIleCysTh 1796
566 TGGCCAAGGAGGCTCCCAATTCCTCAGGCCAGTGTACAGCCCTGATG 615
1796 rAlaLysSerLysSerProAlaTyThrLeuValTrpThrArgLeu.... 1811
616 AATCAGTGAATGGAAAAAGTTACTTGGAACTACTGGATTAATGGAGCA 665
1812 ....HisAsnGlyLysLeuProThrArgAlaMetAspPheAsnGlyIle 1826
666 GTGCTGTGCTACT.....AAGGATGACGGTGCTACT..... 698
1827 LeuThrIleArgAsnValGlnLeuSerAspAlaGlyThrTyValCysTh 1843
699 ....TCAAGGTATTTCACAACTTATGACACGGAATGGTAGATACAGTGA 744
1843 rGlySerAsnMetPheAlaMet.....AspGlnGlyThrAlaThrLeuH 1858
745 AAGTCGGGCTCTGGAGGAGTTAAAGCAGCAGCAGGAGAGTATACCC 794
1858 isValGlnAla..... 1861
795 CAGCAGATGGAGACTGTACATACCTGGCTGGATTGAGAATGATGAAT 844
1862 .....SerGlyThrLeuSerAlaPro.....ValVa 1870
845 ACAATGGAATCCACCAAGACCTGAAATTAATGAAGTATGATGTTCACACA 894
```

```
1870 lSerIleHisProGlnLeu.....ThrValGlnProG 1882
895 AGCAAGTGTGTTTACAGCAACATCTCGGAGGCTCATTTGTGGCTTCT 944
1882 yGlnLeuAlaGluPheArgCysSerAlaThrGlySer..... 1894
945 GATGTCCCAAAATGCTGCCATACCTGATCTC.....TTCCACC 982
1895 .....ProThrProThrLeuGluTrpThrGlyGlyProG 1906
983 TGGCCAAATCACCAGCTGAAGCGCAAAATTCACGGGGCAGTCTCATTA 1032
1906 yGlyGlnLeu...ProAlaLysAlaGlnIleHisGlyGly...IleLeuA 1921
1033 ATCTGACTTGGACAGCTCCTGGGATGATTATGAC..... 1067
1921 rGleuProAlaValGluProThrAspGlnAlaGlnTyLeuCysArgAla 1937
1068 CATGGAACAGCTCACAAATATATCATTCGAATAAGTACAGTATCTTGA 1117
1938 HisSerSerAlaGlyGlnGlnValAlaArg.....AlaValLeuH 1951
1118 TCTCAGACACAAGTTCAATGAATCTCTCAAGTGAAT..... 1154
1951 sValHisGlyGlyGlyProArgValGlnValSerProGluArgThrG 1968
1155 .....ACTACTGCTCTC..... 1166
1968 lnValHisAlaGlyArgThrValArgLeuTyrcysArgAlaAlaGlyVal 1984
1167 .....ATCCCAAAGGA 1177
1985 ProSerAlaThrIleThrTrpArgLysGluGlySerLeuProG 2001
1178 AGCCAACCTGAGGAA.....GTCCTTTGTTTAAACACAGAAAACA 1218
2001 nAlaArgSerGluArgThrAspIleAlaThrLeuLeuIleProAlaIle 2018
1219 TTACTTTTGAATAATGGCAGATCTTTTCATTGCTATTTCAGGCTGTGAT 1268
2018 hrThrAlaAspAlaGlyPheTyrcysValAlaThrSerProAlaGly 2034
1269 AAGTCGATCTGAAATCAGAAATATCCAACTGACAGATATCTTTGTT 1318
2035 ThrAlaGlnAlaArgIleGlnValValValLeuSerAlaSerAlase 2051
1319 TATTCCTCCACAGACTCCGCGACAGACACTAGTCTGATGAACGCTCG 1368
2051 rGlnProPro...ValLysIleGluSerSerProSerValThrGluG 2067
1369 CTCCTTGTCTTAATATCATATCAACAGACCACTCTCGCATTCACATT 1418
2067 ly.....GlnThrLeuAspLeuAsnCysValValAlaGlySerAlaHis 2081
1419 TTAATAATATCTGG 1433
2082 AlaGlnValThrTrp 2086
seq_name: SwissProt_39:YPR4_CAEEL
seq_documentation_block:
ID YPR4_CAEEL STANDARD; PRT; 686 AA.
AC Q20059;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 77.0 KDA TRP-ASP REPEATS CONTAINING PROTEIN F35G12.4 IN
DE CHROMOSOME III.
GN F35G12.4.
OS Caenorhabditis elegans.
CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
CC Rhabditidae; Peloderinae; Caenorhabditis.
```

```

alignment_block:
  US-09-049-696-19 x YPR4_CAEEL  ..
  Align seg 1/1 to: YPR4_CAEEL from: 1 to: 686

160 GAGGAGCTGTCAAAATACAGAGGAGTTTACAGACATATGCTTCAGATCA 109
    |||::: ||::: |||
211 GluLysIleMetLysLeuArgGlyHisThrAspAsnValArgAlaLeuVal 227
    |||::: ||::: |||
110 AGTTTCAGACAATGGCGCTCATGTATGCTTTGGGGCGCTTTTCATCAGGAA 159
    |||::: ||::: |||
227 lValAsnAspGlyThr.....ArgAlaLeuSerAlaGlyS 240
    |||::: ||::: |||
160 ATGAGAGCTGTC.....TCTCAGCGCTCCATCCAG 188
    :: ||::: |||
240 erAspAlaThrIleArgLeuTrpAspIleGlyGlnArgCysIle... 255
    |||::: ||::: |||
189 CTTGAGAGTAAAGGATTACCTCCAGACACCCAGTGGATGAATGGCAC 238
    |||::: ||::: |||
256 .....AlaThrCysIleAlaHisGluGluGlyValTrp.....Th 267
    |||::: ||::: |||
239 AGTCATCGTGGACAGC.....ACCGTGGGAAAGGACA 270
    |||::: ||::: |||
267 rLeuGlnValAspSerSerPheThrThrValTrpSerAlaGlyLysAspL 284
    |||::: ||::: |||
271 CTTTGGTTTCTTTATACC.....TGGACACAGCAGCCTCCCCAAATCCTT 314
    |||::: ||::: |||
284 ysmetValLysThrProLeuTyrrAspPheThrLysSerGlnLeuLeu 300
    |||::: ||::: |||
315 CTCCTGGATCCAGTGGACAGAAGGAGGTGGCTTTGTATGGACAAAAA 364
    |||::: ||::: |||
301 pheLysGluGlnAlaProValLysVal..... 309

```

```

||||| ||||| ||||| ||||| ||||| |||||
539 uillelleCysGluGlyAsnGlyArgProLeuTyrrArgLeuLeuValGlyA 556
1210 CAGAAACATTACTTTTGA...AATGGCACAGATCTTTTTCATTGCTATT 1256
556 spAlaGlyLysGluPheGluAlaAsnGluLeuAlaGlnIleAlaProMet 572
1257 CAGGCTGTTCATAGCTGCATCTGAATCAGAAATATCCAACATTGCACG 1306
573 TrpValIleAspAlaIleGlu...ArgAsnGlnLeuProLysPheAsnLy 588
1307 AGTATCTTTGTTTATT.....CCTCCACAGACTCCGCCAGACACACTA 1350
588 smetProPheTyrrLeuLeuProHisProSerThrAsnProLysGlnProL 605
1351 GTCCTGATGAACGCTGCT 1370
605 ysLysAspArgLeuSerAla 611
seq_name: SwissProt_39:PMPB_CHLMU

```

```

seq_documentation_block:
ID PMPB_CHLMU STANDARD; PRT; 1672 AA.
AC Q9PJY2;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE OUTER MEMBRANE PROTEIN PMPB PRECURSOR (POLYMORPHIC MEMBRANE
DE PROTEIN B).
DE PMPB OR TC0694.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg.
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin N., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39".
PL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC -!- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002338; AAF39510.1; -
DR TIGR; TC0694; -
DR InterPro; IPR003357; OMP.
DR Pfam; PF02385; OMP; 1.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 14 POTENTIAL.
FT CHAIN 15 1672 PROBABLE OUTER MEMBRANE PROTEIN PMPB.
SQ SEQUENCE 1672 AA; 176295 MW; 4ABF190DA4DF8BD6 CRC64;

```

```

alignment_scores:
  Quality: 130.00      Length: 548
  Ratio: 0.570        Gaps: 22
  Percent Similarity: 41.606      Percent Identity: 19.161

```

```

alignment_block:

```

```

US-09-049-696-19 x PMPB_CHLMU
Align seg 1/1 to: PMPB_CHLMU from: 1 to: 1672
9 GTGGCATCATCCACAGACTCGCTTTGGGGCCCTCTGCAGCTCAAGAACT 58
||||| ||||| ||||| ||||| ||||| |||||
263 GlyAlaIleTyrrAlaThrGlyAspLeuThrIleSerAspSerGlnGluI 279
59 A.....GAGGAGCTGTCCAAAATGACAGAGGTTT..... 89
279 eValPheSerValAsnLysAlaSerLysAspGlyGlyAlaIlePheAlaG 296
90 .....CAGACATATGCTTCAGATCAAGTTCAGAACAAAT 122
296 luLysAsnValSerPheGluAsnIleThrThrLeuLysValGlnAsnAsn 312
123 GGCTCATTTGATGCTTTTGGGGCCCTTCATCAGAGAAATGGAGCTGTCTC 172
||||| ||||| ||||| ||||| ||||| |||||
313 .....GlyAlaGluGluLysGlyGlyIleTyrrAl 323
173 TCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAACCTCCAGAACAGCC 222
18 ..... ||||| ||||| ||||| ||||| |||||
323 aSerGlyAspLeuSerIleGlnSerSerLysGlnSerLeuPheAsnSer. 339
223 ACTGGATGAATGGCACAGTTCATCTGGACAGACCCCTGGGAAAGGACACT 272
339 ..... 339
273 TTGTTTCTTATCACCTGGACAGCGCCTCCCAAAATCCTCTCTGGGA 322
340 .....AS 340
323 TCCAGTGACAGACAGCAAGTGGCTTTGTAGTGGACAAAACACCAAAA 372
||||| ||||| ||||| ||||| ||||| |||||
340 nThrSerLysGlnGlyGlyAlaLeuTyrrIleGluGlyAsnValasp 357
373 TCGCTTACCTCCAAAATCCAGGACATTGCTAAGTTGGCAGCTTGAATAC 422
||| ||||| ||||| ||||| ||||| |||||
357 heLysAspLeuGluGluIleArgIle.....LysTyrr 367
423 AGTCTGCAAGCAAGCTCACAAAACCTTGACCTGACTGTACAGTCCGCTGC 472
||| ||||| ||||| ||||| ||||| |||||
368 AsnLysSerGlyThrPheGluThrLysLysValThrLeuSerLeuProGl 384
473 GTCCAATGTACCTGCTCCCAATTCAGTACTTCCAAACCAAGCAAGG 522
||||| ||||| ||||| ||||| ||||| |||||
384 uAlaGlnThrAsnLysSerSerValThrAlaAlaSerGlnSerGlyProA 401
523 ACACGACGAAATCCCGACCCCTCTGTA.....GTT 554
||| ||||| ||||| ||||| ||||| |||||
401 snThrThrProThrProThrProValThrAlaLysGlyGlyLeu 417
555 TATGCAATATTCGCCAAGAGGCTCCCAATTCAGGGCCAGTGTCTAC 604
||||| ||||| ||||| ||||| ||||| |||||
418 TyrThrGluLysAsnLeuSerIleSer.....AsnIleThr 429
605 AGCCCTGATTGAATCAGTGAATGGAATAACAGTACCTTGGAACTACTGG 654
||||| ||||| ||||| ||||| ||||| |||||
429 rGlyIleIleGluIleThrAsnAsnLysAlaThr.....A 441
655 ATAATGGACAGCTGCT..... 671
||| ||||| ||||| ||||| ||||| |||||
441 spValGlyGlyAlaTyrrValLysGlyThrLeuThrCysLysAspSer 457
672 .....GATGCTACTAAGGATGACCGTGT 694
453 HisArgLeuGlnPheGlnLysAsnSerSerGluLysLysGlyGlyLe 474
695 CTACTCAAGGATTTTCAACAATTATGACACAGAAATGTTAGATACAGTGA 744
||||| ||||| ||||| ||||| ||||| |||||
474 uTyrr.....ThrGluAspThr.....IleT 481
745 AAGTGGGGCTCTGGGAGGAGTTTAACCGACCGGAGAGTGATACCC 794

```

OS	Mus musculus (Mouse).	
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
CC	NCBI_TaxID=10090;	
CC	[1]	
RN	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.	
RP	TISSUE=Brain;	
RC	MEDLINE=97407661; PubMed=9264410;	
RA	Keeling S.L., Gad J.M., Cooper H.M.;	
RT	"Mouse neogenin, a DCC-like molecule, has four splice variants and is expressed widely in the adult mouse and during embryogenesis.";	
RL	OncoGene 15:691-700(1997).	
CC	FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.	
CC	SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.	
CC	ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS: 1 (SHOWN HERE), 2, 3, 4 AND 5; ARE PRODUCED BY ALTERNATIVE SPLICING. THE EXPRESSION OF ISOFORMS 3, 4 AND 5 ARE DEVELOPMENTALLY REGULATED.	
CC	TISSUE SPECIFICITY: WIDELY EXPRESSED.	
CC	DEVELOPMENTAL STAGE: EXPRESSED UBIQUITOUSLY THROUGHOUT THE MID TO LATE STAGES OF GESTATION AND IN ADULT TISSUES. STRONG EXPRESSION IS OBSERVED IN THE VENTRAL REGION OF THE VENTRICULAR ZONE OF THE E15.5 MOUSE NEURAL TUBE, AS WELL AS IN THE VENTRICULAR ZONES OF THE MESENCEPHALON AND RHOMBENCEPHALON. ISOFORMS 3 AND 4 ARE EXPRESSED AT HIGHER LEVEL COMPARED TO OTHER ISOFORMS BETWEEN E11.5 AND E16.5.	
CC	TISSUE SPECIFICITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.	
CC	SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.	
CC	SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO TUMOR SUPPRESSOR PROTEIN DCC.	
CC	-----	
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; Y09535; CAA70727.1; .	
DR	HSSP; P02751; ITTG.	
DR	MGI; MGI:1097159; Neol.	
DR	InterPro; IPR001777; FN.III.	
DR	InterPro; IPR003006; Ig_MHC.	
DR	InterPro; IPR003598; Ig_C2.	
DR	Ffam; PF00041; fn3; 6.	
DR	Ffam; PF00047; ig; 4.	
DR	PRINTS; PR00014; FNTYPEIII.	
DR	SMART; SM00408; IGC2; 6.	
DR	SMART; SM00408; IGC2; 4.	
KW	Transmembrane; Immunoglobulin domain; Glycoprotein; Signal;	
KW	Alternative splicing.	
FT	SIGNAL 1 36	POTENTIAL.
FT	CHAIN 37 1493	NEOGENIN.
FT	DOMAIN 37 1136	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 1137 1157	POTENTIAL.
FT	DOMAIN 1158 1493	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 78 147	IG-LIKE C2-TYPE DOMAIN 1.
FT	DOMAIN 177 239	IG-LIKE C2-TYPE DOMAIN 2.
FT	DOMAIN 274 338	IG-LIKE C2-TYPE DOMAIN 3.
FT	DOMAIN 366 428	IG-LIKE C2-TYPE DOMAIN 4.
FT	DOMAIN 467 564	FIBRONECTIN TYPE-III 1.
FT	DOMAIN 567 660	FIBRONECTIN TYPE-III 2.
FT	DOMAIN 661 760	FIBRONECTIN TYPE-III 3.
FT	DOMAIN 766 860	FIBRONECTIN TYPE-III 4.
FT	DOMAIN 861 981	FIBRONECTIN TYPE-III 5.
FT	DOMAIN 982 1083	FIBRONECTIN TYPE-III 6.
FT	DOMAIN 1149 1153	POLY-VAL.
FT	DOMAIN 84 84	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 221 221	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 337 337	N-LINKED (GLCNAC. . .) (POTENTIAL).


```

FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 746 746 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 940 940 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 442 461 MISSING (IN ISOFORM 2).
FT VARSPLIC 863 878 MISSING (IN ISOFORM 3).
FT VARSPLIC 1086 1096 MISSING (IN ISOFORM 4).
FT VARSPLIC 1279 1331 MISSING (IN ISOFORM 5).
SQ SEQUENCE 1493 AA; 163159 MW; 441DE919D5E17C0E CRC64;

```

```

alignment_scores:
  Quality: 124.50      Length: 539
  Ratio: 0.546         Gaps: 31
  Percent Similarity: 42.301      Percent Identity: 21.707

```

alignment_block:

US-09-049-696-19 x NEO1_MOUSE

Align seg 1/1 to: NEO1_MOUSE from: 1 to: 1493

```

162 GGAGCTGTCTCTCAGCGCTCCAGCTGAGAGTAAAGGATTAACCCCT 211
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 GlyThrIleValSerArgThrAlaLysLeuThrValAlaGlyLeuProAr 164
212 CCAGAACAGCCAG.....TGGATGAATGGCAGCAGTGA 243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
164 gPheThrSerGlnProGluProSerSerValTyValGlyAsnSerAlaI 181
244 TCCTGGACAGCACCCTGGGAAGGACACATTGTTCTTATCACCCTGGACA 293
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 leuAsnGlnProLeuLeuLeuAspValProPheValArgTrpGlu 197
294 AGCAGCCTCCCAATCTCTCTGGGAT.....CCGAG 328
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 GlnAsnArgGlnProLeuLeuLeuAspAspArgIleValLysLeuProSe 214
329 TGA.....CAGAGCAAGTGGCTTT.... 350
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 rGlyThrLeuValIleSerAsnAlaThrGluGlyAspGlyGlyLeuTyra 231
351 .....GTAGTGACAAACACCAAAATGGCCTACCTCCCAATCCAGGC 395
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
231 rGcysIleValGluSerGly..... 237
396 ATTGCTAAGTTGGCACTTGGAAATACAGT..... 425
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 .....GlyProLysPheSerAspGluAlaGluLeuLysVa 250
426 .CTGCAGCAAGCTCAACAACTTGACCTGACTGTACGTCCCGTGGCGT 474
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
250 lLeuGlnAspProGluGluIleValAspLeuPheLeuMetArgProS 267
475 CC.....AATGCTACCTGCTCCCAATTACA 500
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
267 eSerMetLysValThrGlyGlnSerAlaValLeuProCysVal... 282
501 GTGACTTCCAAACGAAACAGGACACCAATATCCCGAGCCCTCTGGT 550
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
283 .....ValSerGlyLeuProAlaProValVa 291
551 AGTTTATGAATATTGGCAGAGGACCTCCCAATCTCAGGGCCAGTG 600
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 lArgTrpMetLysAsnGluGlu..... 298
601 TCACACCCCTGATTGAATCAGTGAATGGAACACAGTACCTTGGAACTA 650
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
299 .....ValLeuAspThrGluSerSerGlyArgLeuValLeuLeu..... 311
651 CTGGATAATGAGCAGGT.....GCTGATGCTACTAAGGA 685
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
312 .....AlaGlyGlyCysLeuGluIleSerAspValThrGluAs 324

```

```

686 TGACGGTGTCTACTCAAGGTATTTCACAACTTATGACACGAATGGTAGAT 735
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
324 pAspAla.....GlyThrTyPheCysIleAlaAsp...AsnGlyAsnL 338
736 ACAGTGTAAAAGTGGCGCTCTGGAGGAGTTAACCGCAGCCAGACGAGA 785
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
338 yThrValGluAlaGlnAlaGluLeuThrVal.....Gln 349
786 GTGATACCC.....CAGCAGAGTGGAGCAGCTGTAC..... 815
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
350 ValProGlyPheLeuLysGlnProAlaAsnIleTyAlaHisGluSe 366
816 .....ATACCTGGC..... 824
366 rMetAspIleValPheGluCysGluValThrGlyLysProThrProThrV 383
825 .....TGGATTGAGAAATGATGAAATCAATGGAATCCACCAAGACCTGAA 869
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
383 alLysTrpValLysAsnGlyAspValValIleProSerAspAsnPheLys 399
870 ATTATAAGGATGATGTTCAACACACAAAGTGTGTTTTCAGCAGAACATC 919
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
400 IleValLysGlu.....HisAsnLeuGlnValLeuGlyLeuValLysSe 414
920 CTCGGGAGGCTCATTT.....GTGGCTTCTGATGTCCCAAAATGCT. 959
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 rAspGluGlyPheTyGlnCysIleAlaGluAsnAspValGlyAsnAlaG 431
960 .....CCCATACCTGAT 971
431 lAlaGlyAlaGlnLeuIleLeuGluHisAspValAlaIleProThr 447
972 CTCTTCCACCTGGCCAAATCACC..... 995
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
449 Leu...ProProThrSerLeuThrSerAlaThrThrAspHisLeuAlaPr 463
996 .....GACCTGAAGCGGAAA 1011
463 oAlaThrThrGlyProLeuProSerAlaProArgAspValValAlaSerL 480
1012 TTCACGGGGCAGCTCATTAATCTGACTTGGACAGCTCCTGGGGATGAT 1061
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
480 euValSerThrArgPheIleLeuThrTrpArgThrProAlaSerAsp 496
1062 TATGACCATGGAACAGCTCACAGTATATATTCGAATAAGTACAAGTAT 1111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
497 ...ProHisGlyAspAsnLeuThrTySerValPheTyThrLysGluG 512
1112 TCTTGATCTCAGACACAAGTTCAATGAATCT.....CTTCAAG 1149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
512 yValAspArgGluArgValGluAsnThrSerGlnProGlyGluMetGlnV 529
1150 TGAATACTACTGCTCTCATCCCAAGGAAGCACTCTGAGGAAGTCTTT 1199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
529 alThrIleGlnAsnLeuMetPro.....AlaThrValTy 540
1200 TTGTTTAAACACGAAAAACATTACTTTTGAATAATGGCACAGATCTTTTCAT 1249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 lIlePheLys..... 543
1250 TGCTATTACGGCTTTCATAGGTTCGATCTGAAATCAGAATATCCAACA 1299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
544 ...ValMetAlaGlnAsnLysHisGly...SerGlyGluSerSerAlaP 558
1300 TTCCACAGTATCTTTCTTTATTCCTCCACAGACTCCGCCAGACACACT 1349
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
558 roLeuArgVal.....GluThrGlnProGluValGln 568
1350 AGTCCTGATGAAACGTCTGCTCCTTGTCTCAATATTCATATCAACAGCAC 1399
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
569 LeuPro.....GlyProAlaProAsnIleArgAlaTyAlaTh 581

```


768	sasnValThrGlyAspLeuProAsnIleTrpLeuGlnTyrGlyInPheL	785	tyrosine phosphatases.;
910	CGAGAATCCTCGGGAGGCTCATTTGTGCTTCT...GATGTCACAAAT	956	EMBL J. 9:3241-3252(1990).
785	ysLeuLysValSerGlyGlyAsnGlyThrTyrSerTrpHisSerGluAsn	801	CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
957	GCTCCCATACTGATCTCTCCACCTGGCCAAATCAGCAGCTGAAGC	1006	PROTEIN TYROSINE + ORTHOPHOSPHATE.
802	ThrAsnIleAlaThrValAspGluSerGlyLysValThr...LeuLysG	817	SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN
1007	GGAATTCAGGGGCGGCTCATCTAATCTGCTGCTGGACAGCTCTGGGG	1056	ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS ARE FOUND IN DIFFERENT
817	YLys...GlyThrAlaValIleAsnValThr.....SerGlyA	829	TISSUES DUE TO ALTERNATIVE SPLICING.
1057	ATGATTATGACCATGGA.....ACAGCTCACAGTATATCATTCGA	1097	PTM: A CLEAVAGE OCCURS THAT SEPARATES THE EXTRACELLULAR DOMAIN
829	spLysGlnThrValSerTyrThrIleLysAlaProAsnTyrMetIleArg	845	FROM THE TRANSMEMBRANE SEGMENT.
1098	ATAAGTACAAGTATCTTTGATCTCAGAGCAAGTTCAAGTCTCTTCA	1147	SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
846	ValGly.....AsnLysAlaSerTy	852	SIMILARITY: CONTAINS 8 FIBRONECTIN TYPE III-LIKE DOMAINS.
1148	AGTGAATACTGCTCTCATCCCAAGGAGCAACTCTGAGGAAGTCT	1197	SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
852	rAlaAsnAlaMetSerPheCysGlyAsnLeuLeuProSerSerGlnThrV	869	THIS SWISS-PROT entry is copyright. It is produced through a collaboration
1198	TTTTGTTTAAACCAAGAAACATTACTTTT...GAAATGGCAGCATCTT	1244	between the Swiss Institute of Bioinformatics and the EMBL outstation
869	allLeuSerAsnValTyrAsnSerTrpGlyProAlaAsnGlyTyrAspHis	885	the European Bioinformatics Institute. There are no restrictions on its
1245	TTCATGCTATTCAGCTGTGTGATGAAGTGTGATCTGAAATCAGAAATATC	1294	use by non-profit institutions as long as its content is in no way
886	TyrArgSerMetGlnSerIleThrAlaTrpIleThrGlnThrGluAlaAs	902	modified and this statement is not removed. Usage by and for commercial
1295	CAACATTCACGAGTATCTTTGTTTATTCTCTCCACAGACTCCGCCAGACA	1344	entities requires a license agreement (See http://www.isb-sib.ch/)
902	plysIleSerGlyValSerThrThrTyrAspLeuIleThr.....GlnA	917	or send an email to license@isb-sib.ch .
1345	CACCTAGTCTGTGAACAGCTGTCTGCTTCTTCTATATATTCAT	1388	EMBL: L38929; AAC41749.1; ..
917	snProHisLysAsp...ValSerLeuAsnAlaProAsnValTyr	930	EMBL: Y54133; CAA38068.1; ..
seq_name:	SwissProt_39:PTPD_HUMAN		PIR: S12052; S12052.
seq_documentation_block:			HSSP: PI8052; LYFO.
ID	PTPD_HUMAN	STANDARD;	MIM: 601598; ..
AC	P23468;	PRT;	InterPro: IPR001777; FN.III.
DT	01-NOV-1991 (Rel. 20, Created)		InterPro: IPR003006; Ig_MHC.
DT	20-OCT-1996 (Rel. 34, Last sequence update)		InterPro: IPR003598; Ig_C2.
DE	PROTEIN-TYROSINE PHOSPHATASE DELTA PRECURSOR (EC 3.1.3.48) (R-PTP-		InterPro: IPR000367; TYR_phosphatase.
DE	DELTA).		InterPro: IPR000242; Tyr_prot_phphatase.
GN	PTPDR.		Pfam: PF00041; fn3; 8.
OS	Homo sapiens (Human).		Pfam: PF00102; Y_phosphatase; 2.
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		PRINTS: PR00014; ENTPELII
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		PRINTS: PR00700; PRYPHPHTASE.
OX	NCBI_TaxID=9606;		SMART: SM00060; FN3; 8.
RP	[1]		SMART: SM00194; PTPC; 2.
RP	SEQUENCE FROM N.A., AND MUTAGENESIS OF ARG-1178.		SMART: SM00408; IGC2; 3.
RX	MEDLINE=95204468; PubMed=7896816;		PROSITE: PS00383; TYR_PHOSPHATASE_1; 2.
RA	Pulido R., Krueger N.X., Serra-Pages C., Saito H., Streuli M.;		PROSITE: PS50056; TYR_PHOSPHATASE_2; 2.
RT	"Molecular characterization of the human transmembrane protein-		PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 2.
RT	tyrosine phosphatase delta. Evidence for tissue-specific expression of		Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane; Duplication;
RT	alternative human transmembrane protein-tyrosine phosphatase delta		Immunoglobulin domain; Alternative splicing.
RT	isoforms.;"		POTENTIAL.
RL	J. Biol. Chem. 270:6722-6728(1995).		CHAIN 1 20
RL	[2]		DOMAIN 21 1912
RP	SEQUENCE OF 390-1912 FROM N.A.		DOMAIN 21 1265
RP	TISSUE-Placenta;		TRANSMEM 1286 1290
RX	MEDLINE=91006018; PubMed=2170109;		DOMAIN 1291 1912
RA	Krueger N.X., Streuli M., Saito H.;		DOMAIN 23 115
RT	"Structural diversity and evolution of human receptor-like protein		DOMAIN 118 225
RT			DOMAIN 232 318
			DOMAIN 320 414
			DOMAIN 417 513
			DOMAIN 516 606
			DOMAIN 609 708
			DOMAIN 711 822
			DOMAIN 825 916
			DOMAIN 918 1017
			DOMAIN 1020 1137
			DOMAIN 1137 1618
			DOMAIN 1619 1912
			ACT_SITE 1553 1553
			ACT_SITE 1844 1844
			STATE 1175 1178
			CARBOHYD 254 254
			CARBOHYD 299 299
			CARBOHYD 724 724
			CARBOHYD 832 832
			CARBOHYD 181 189
			VARSPIC 226 229
			VARSPIC 775

Tue Apr 2 09:39:54 2002

us-09-049-696-19.rsp

FT MUTAGEN 1178 1178 R->A: 2.5-FOLD REDUCTION IN CLEAVAGE.
SQ SEQUENCE 1912 AA; 214759 MW; 3A88CBDC32182E26 CRC64;

alignment_scores:
Quality: 124.00 Length: 361
Ratio: 0.752 Gaps: 18
Percent Similarity: 45.706 Percent Identity: 22.715

alignment_block:

US-09-049-696-19 x PTPD_HUMAN ..

Align seg 1/1 to: PTPD_HUMAN from: 1 to: 1912

```

246 GTGACAGCACCGTGGGAAAGACACATTTCTTCTATCACCTGGACAAC 295
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
616 ILeSerCysThrProSerThrSerThrSerThrSerThrSerThrSer 630
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
296 GCAGCCTCCCAATCCTCTCTGGGATCCAGTGGACAGACAAGAGGTG 345
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
631 GlnProProVal.....GluLysGlnAsnG 640
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
346 GCTTTGAGTGGACAAAACACCAAAATGCCCTACCTC..... 383
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
640 LyllelleThrGluThrSerIleLysThrAlaValAspGlyGluAsp 656
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
384 .....CAATCCAGGACATGCTAAGTTGGCAGCTTGGAAATA 421
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
657 AspLysProHisGluLeuGlyLeuIleProSer....AspThrThrLys 672
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
422 CAGTCTGCAAGCAAGCTCACAA.....ACCTTGACCTGACTGTACGT 465
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
672 rLeuLeuGluGlnLeuGluLysThrThrGluThrArgIleThrValThrA 689
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
466 CCGTGCCTCCAACTGCTACCTGCT...CCAATTACAGTACCTCCAAA 512
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
689 laHisThrAspValGlyProGlySerLeuSerValLeuIleArg 705
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
513 AGCAACAGGACACAGCAAAATCCACAGCCCTCTGTAGTTATGCAAA 562
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
706 ThrAsnGluAspValProSerGlyProProArgLysValGluValGlu 722
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
563 TATTCGCCAAGGAGGC.....TCCCAATT..... 587
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
722 aValAsnSerThrSerValLysValSerThrArgSerProValProAsnL 739
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
588 .....CTCAGGCGCAGTGCACAGCCCTGATTGAATCAGTG 623
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
739 ySglnHisGlyGlnIleArgGlyThrGlnValHisThrValArgMetGlu 755
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
624 AATGAAAAACAGTTACCTTGGAACTACTGGATAATGAGCAGGTGCTGA 673
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
756 AsnGlyGluProLysGlyGlnProMetLeuLysAspValMetLeuAlaAs 772
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
674 TGCT.....ACTAAGGATGACGCTGC..... 695
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
772 pAlaGlnTrpGluPheAspThrThrGluHisaspMetIleIleSerG 789
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
696 .....TACTCAAGTATTTCACAACTTATGACAGC 725
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
789 lLeuGlnProGluThrSerThrSerLeuThrValThrAlaThrThrThr 805
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
726 AATGGT.....AGATACAGTGTAAA.....GTGCGGCTCTGGAGG 763
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
806 LysGlyAspGlyAlaArgSerLysProLysLeuValSerThrThrGlyAl 822
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
764 AGTTAAGCGCCAGCAGGAGAGTGATA.....CCCCAGCAGATGGAG 807
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
822 aValProGlyLysProArgLeuValIleAsnHisThrGlnMetAsnThrA 839
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
808 CACTGTACATACCTGGCTGATTGAGATGAAATACAAATGGAATCCA 857
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
839 laLeu.....IleGlnTrpHisPro 845

```

```

858 CCA.....AGACCTGAAATTAATAA 877
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
846 ProValAspThrPheGlyProLeuGlnGlyTyrArgLeuLysPheGlyAr 862
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
878 GGATGATGTTCAA.....CACAAACAAGTGT 903
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
862 gLysAspMetGluProLeuThrThrLeuGluPheSerGluLysGluAspH 879
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
904 GTTTCAGCAGAACATCC.....TCGGGAGGCTCATTTGTG..... 938
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
879 isPheThrAlaThrAspIleHisLysGlyAlaSerTyrValPheArgLeu 895
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
939 .....GCTTCTGATGTCCTCAATGCTCC 961
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
896 SerAlaArgAsnLysValGlyPheGlyGluMetValLysGluIleSe 912
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
962 CATACCTGATCTCTCCACCTGCCAAATCACCGACCTGAAGCGGAAA 1011
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
912 fileProGluGluValProThrGlyPheProGlnAsnLeuHisSerGlu 929
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1012 TTCACGGGGCAGTCTCAATTAATGACTTGGACAGCTCTCGGGGATCAT 1061
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
929 lYThrThrSerThrSerValGlnLeuSerTrpGlnProProValLeuAla 945
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1062 TATGACCATGGACAGCTCACAAATGATATCAT 1094
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
946 GluArgAsnGlyIleIleThrLysThrThrLeu 956
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
seq_names: SwissProt_39:MAC1_YEAST
seq_documentation_block:
ID MAC1_YEAST STANDARD; PRT; 417 AA.
AC P35192;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE METAL BINDING ACTIVATOR 1.
GN MAC1 OR CUA1 OR YMR021C OR YMR711.11C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
CX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94085378; PubMed=8262047;
RA Jungmann J., Reins H.-A., Lee J., Romeo A., Hassett R., Kosman D.,
RT "MAC1, a nuclear regulatory protein related to Cu-dependent
FT transcription factors is involved in Cu/Fe utilization and stress
RL resistance in yeast."
RL EMBO J. 12:5051-5056(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Lye G., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
FL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REGULATORY PROTEIN INVOLVED IN CU/FE UTILIZATION AND
CC STRESS RESISTANCE. INVOLVED IN BASAL LEVEL TRANSCRIPTION OF FRE1
CC AND H(2)O(2)-INDUCED TRANSCRIPTION OF CTT1.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: CONTAINS A "COPPER-FIST" DNA-BINDING DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X74551; CAA52645.1; -.
CC EMBL; Z49211; CAA89124.1; -.
DR

```

DR PIR: S41687.
 DR TRANSFAC; T01265; ..
 DR SGD; S0004623; MAC1.
 DR InterPro; IPR001083; Copper-fist.
 DR Pfam; PF00649; Copper-fist; 1.
 DR PRINTS; PD00617; COPPERFIST.
 DR ProDom; PD009050; Copper-fist; 1.
 DR SMART; SM00412; Cu_FIST; 1.
 DR PROSITE; PS01119; COPPER_FIST_1; 1.
 DR PROSITE; PS00073; COPPER_FIST_2; 1.
 KW Transcription regulation; DNA-binding; Activator; Metal-binding;
 KW Copper; Trans-acting factor; Nuclear protein; Repeat.
 FT DNA_BIND 1 40
 FT DOMAIN 264 337
 FT REPEAT 264 279
 FT REPEAT 322 337
 FT MUTAGEN 279 279 H->Q: GAIN OF FUNCTION.
 FT CONFLICT 198 198 P -> A (IN REF. 1).
 SQ SEQUENCE 417 AA; 46516 MW; 2D665168F04BEC5D CRC64;

alignment_scores:

Quality: 120.50 Length: 346
 Ratio: 0.803 Gaps: 15
 Percent Similarity: 43.353 Percent Identity: 19.942

alignment_block:

US-09-049-696-19 x MAC1_YEAST ..

Align seg 1/1 to: MAC1_YEAST from: 1 to: 417

537 CCAGCCCTCTGGTAGTT..... 554
 |||||.....
 39 ProSerProMetAlaIleArgAspAlaIleLeuValAspSerThrserG1 55
 555TATGCAAAATATCCCAA..... 572
 ||| ||| |||
 55 nSerThrGluTyroGluAsnGlyAlaGlnIleGluGlyAspCysSerA 72
 573GGAGCCTCCCAATCTC.....AGGCCAGTGTACAGCCCTG 611
 : : : : : |||||
 72 laMetAsnGlnProIleLeuPheValArgAlaSerAlaValArgLys 88
 612 ATTGAATCAGTGAATGAAACAGTTACCTTGGAACTACTGATATGG 661
 : : : : : |||||
 89 AlaArgMetIleAsnGlyLysLeuHisile.....LeuMetGluGluG1 103
 662 AGCAGGTGCTGCTACTAAGATGACGCTGTCTAC.....T 699
 | ||| ||||| : : : : :
 103 yPheThrAlaHisGluProLysAspIleSerThrPheThrAspAspGlyA 120
 700 CAAGGTATTTACAACTATTATGACAGNAATGGTAGATACAGTGTAAAGT 749
 : : : : : ||||| : : : : :
 120 snLysTyrlleThrGluThrGluPheLeuArgLysHisSerProLysAla 136
 750 CGGCTCTGGAGGAGTACCGAGCCAGGAGGAGATACCCAGCA 799
 ||| ||| : : : : :
 137 ProAlaThrGlyThrIleSer..... 143
 800 GAGTGGAGCAGCTGTACATACATCTGGCTGGATTGAGAAATGATAACAAT 849
 ||| ||| : : : : :
 144ProAspSerThrLysSerSerSerSerSerg 154
 850 GGAATCCCAACAGCCTGAAATTAATAGGATGATGTTCACACACAGCAA 899
 : : : : : ||||| : : : : :
 154 LuLysLysGluArgSerArgLeuGlnGlnGluProIleArgHisPheSer 170
 900 GTGTGTTTACAGCAACATCCTCGGAGGCTCATTTGTGGCTCTGTGATG 949
 ||| : : : : : ||||| : : : : :
 171 AsnCysCysLysLysAspLysSerGlnAsnProAlaSerAsnGlyLysTh 187
 950 CCAAAATGCTCCCATACCTGATCTCTTCCCACTGGCCAAATCACCAGC 999

187 rAsnLysAlaProSerAspIlePheThrPro..... 198
 1000 TGAAGCGGAAATTCACGGGGCAGTCTCTAATTAATCTGACTTGGACAGCT 1049
 198 198
 1050 CCTGGGGATGATTATGACCATGGAACAGCTCACAAAGTATATCATTCGAAT 1099
 199TyrGlySer..... 201
 1100 AAGTACAAGTATCTCTGATCTCAGAGACAAAGTCAATGAATCTCTCAAG 1149
 202LeuGluSerThrSerAlaPheAsnAspIleLeuGlnG 214
 1150 TGAATACTACTGCTCTCATCCCAAGGAAGCAACTCTGAGGAAGTCTTT 1199
 ||| : : : : : |||||
 214 luAsnTyraSerSerValProGlyAlaHisAspSerSerGluThr... 229
 1200 TTGTTTAAACCAAGAAACATTAATCTTTTGAANAATGGCACAGATCTTTTCAT 1249
 : : : : : |||||
 230 ...LeuThrProGlnSerThrThr..... 236
 1250 TGCTATTTCAGGCTGTGATGAAGTTCGATCTCAATCAGAAATATATCCAACA 1299
 : : : : : ||||| : : : : :
 237 .ThrIleAlaAlaProHisSerSerAspValAlaSerLysValGluVal 253
 1300 TTGCACGAGTATCTTTGTTTATTCCTCCACAGACTCCGCCAGAGACACCT 1349
 : : : : : |||||
 253 eutThrHisLysGlyIlePheLeuSerThrGlnCys..... 264
 1350 AGTCCTGATGAAACGCTGCTCTTGTCTTAAT.....ATTCAT..... 1388
 ||| : : : : : |||||
 265 SerCysGluAspGluSerCysProCysValAsnCysLeuIleHisArgSe 281
 1389ATCAACAGACACCACTT.....CCTGGCATTACATTTAA 1422
 : : : : : |||||
 281 rGluGluGluLeuAsnSerTyrlleGlnGlnSerGlyValProLeuThrA 298
 1423 AAATTATGTGGAAGTGGATAGGAGAACTGCAGCTGTCA 1460
 : : : : : |||||
 298 sn.....IleGlyGluAlaGlnIleThr 305

seq_name: SwissProt_39:SPG7_DICDI

seq_documentation_block:

ID SPG7_DICDI STANDARD; PRT; 532 AA.
 AC P22698;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE SPORE GERMINATION PROTEIN 270-11.
 GN CELB OR GERE.
 OS Dictyostelium discoideum (Slime mold).
 CC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91002566; PubMed=1976383;
 RA Giorda R., Ohmachi T., Shaw D.R., Ennis H.L.;
 RT "A shared internal threonine-glutamic acid-threonine-proline repeat
 RT defines a family of Dictyostelium discoideum spore germination
 RT specific proteins.";
 RL Biochemistry 29:7264-7269(1990).
 CC -!- DEVELOPMENTAL STAGE: FOUND PREDOMINANTLY DURING EARLY STAGES OF
 CC -!- SPORE GERMINATION.
 CC -!- SIMILARITY: A SHARED INTERNAL THR-GLU-THR-PRO REPEAT DEFINES A
 CC FAMILY OF D.DISCOIDEUM SPORE GERMINATION SPECIFIC PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way


```

1032 .....AATCTGACCTGGACAGCTCCTCGGATGATT 1062
      |||||
958 aserAsnAlaThrArgAsnAsnLeuGluValThrLeuAlaAsnGlySerV 975
      |||||
1063 ATGACCATGGACAGCTCACAAATATATCATTCGAATAAGTACAAGTATT 1112
      |||||
975 alaSpArgGly...AlaTrpIysTrpIysLeuArgAsnValAsnGlyArg 990
      |||||
1113 CTTGATCTCAGA.....GACAAGTTCAATGAATCTCTTCAAGT 1150
      |||||
991 TyrAspLeuTyrAsnProGluValGluLysArgAsnGlnThr.....Va 1005
      |||||
1151 GAATACTACTGCTCTCAATCCCAAGGAACCACTCTGAGGAAGTCTTTT 1200
      |||||
1005 lAspThrThr..... 1008
      |||||
1201 TGTTTAAACCAAGAAACATTACTTTTGAATATGGACAGATCTTTTCAAT 1250
      |||||
1009 .....AsnIleThrThrProAsnAsp..... 1015
      |||||
1251 GCTATTACAGCTGTGATAAGTCGATCGAATCAGAAATATCAACAT 1300
      |||||
1016 ...IleGlnAla...AspAlaProSerAlaGlnSerAsnGluGluI 1030
      |||||
1301 TGCAGGATATCTTTGTTTATTCTCCACAGACTCCG..... 1337
      |||||
1030 eAlaArgValGluThrProValProProAlaProAlaThrGluSerA 1047
      |||||
1338 .....CCAGACAGACCTAGTCTCTGATGAACAGCTCTGCT 1370
      |||||
1047 laIleAlaSerGluGlnProGluThr...ArgProAlaGluThrAlaGln 1062
      |||||
1371 CCT 1373
1063 Pro 1063

seq_name: SwissProt_39:DCC_HUMAN
seq_documentation_block:
ID_DCC_HUMAN STANDARD; PRT: 1447 AA.
AC P43146;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TUMOR SUPPRESSOR PROTEIN DCC PRECURSOR (COLORECTAL CANCER SUPPRESSOR).
GN DCC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95011532; PubMed=7926722;
RA Hedrick L., Cho K.R., Fearon E.R., Wu T.-C., Kinzler K.W.,
RA Vogelstein B.;
RT "The DCC gene product in cellular differentiation and colorectal
RT tumorigenesis.";
RL Genes Dev. 8:1174-1183(1994).
RN [2]
RP SEQUENCE OF 1-750 FROM N.A.
RX MEDLINE=90100559; PubMed=2294591;
RA Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,
RA Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,
RA Vogelstein B.;
RT "Identification of a chromosome 18q gene that is altered in
RT colorectal cancers.";
RL Science 247:49-56(1990).
RN [3]
RP SEQUENCE OF 107-472 FROM N.A. (SCRAMBELD EXONS).
RX MEDLINE=91121517; PubMed=1991322;
RA Nigro J.M., Cho K.R., Fearon E.R., Kern S.E., Ruppert J.M.,
RA Oliner J.D., Kinzler K.W., Vogelstein B.;

```

"Scrambled exons.";
 RT Cell 64:607-613(1991).
 RN [4]
 RP GENE STRUCTURE, AND VARIANTS CARCINOMA HIS-1375.
 RX MEDLINE=94245241; PubMed=8188295;
 RA Cho K.R., Oliner J.D., Simons J.W., Hedrick L., Fearon E.R.,
 RA Preisinger A.C., Hedge P., Silverman G.A., Vogelstein B.;
 RT "The DCC gene: structural analysis and mutations in colorectal
 RT carcinomas.";
 RL Genomics 19:525-531(1994).
 RN [5]
 RP VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.
 RX MEDLINE=94243823; PubMed=8187090;
 RA Miyake S., Nagai K., Yoshino K., Oto M., Endo M., Yuasa Y.;
 RT "Point mutations and allelic deletion of tumor suppressor gene DCC in
 RT human esophageal squamous cell carcinomas and their relation to
 RT metastasis.";
 RL Cancer Res. 54:3007-3010(1994).
 CC - FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
 CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC - TISSUE SPECIFICITY: FOUND IN AXONS OF THE CENTRAL AND PERIPHERAL
 CC NERVOUS SYSTEM AND IN DIFFERENTIATED CELL TYPES OF THE INTESTINE.
 CC DISEASE: COLORECTAL TUMORS THAT LOST THEIR CAPACITY TO
 CC DIFFERENTIATE INTO MUCUS PRODUCING CELLS UNIFORMLY LACK DCC
 CC EXPRESSION. INACTIVATION OF DCC DUE TO ALLELIC DELETION AND/OR
 CC POINT MUTATIONS MAY CAUSE BOTH LYMPHATIC AND HEMATOGENOUS
 CC METASTASIS OF OESOPHAGEAL SQUAMOUS CELL CARCINOMAS.
 CC - SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC - SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X76132; CAA53735.1; -
 CC EMBL: M32292; AAA35751.1; -
 CC EMBL: M32286; AAA52174.1; -
 CC EMBL: M32288; AAA52175.1; ALT_SEQ.
 CC EMBL: M32290; AAA52176.1; -
 CC EMBL: M63696; AAA52177.1; -
 CC EMBL: M63700; AAA52178.1; -
 CC EMBL: M63702; AAA52179.1; -
 CC EMBL: M63718; AAA52180.1; -
 CC EMBL: M63698; AAA52181.1; -
 CC PIR: A54100; A54100.
 CC PIR: A40098; A40098.
 CC PIR: A38442; A38442.
 CC PSEP: P56276; 1TLK.
 CC KIM: 120470; -
 CC InterPro: IPR001777; FN_III.
 CC InterPro: IPR003006; Ig_MHC.
 CC InterPro: IPR003598; Ig_c2.
 CC InterPro: IPR003600; Ig_like.
 CC Pfam: PF00041; fn3; 6.
 CC Pfam: PF00047; ig; 4.
 CC PRINTS: PR00014; ENTPEIII.
 CC SMART: SM00060; FN3; 6.
 CC SMART: SM00408; IgC2; 3.
 CC SMART: SM00410; Ig_like; 2.
 CC Givcoprotein; Immunoglobulin domain; Transmembrane; Signal;
 KW Anti-oncogene; Disease mutation; Polymorphism.
 FT SIGNAL 1 25
 FT CHAIN 26 1447
 FT DOMAIN 26 1097
 FT TRANSMEM 1098 1122
 FT DOMAIN 1123 1447
 FT DOMAIN 54 124
 FT DOMAIN 154 219
 FT DOMAIN 254 317


```

FT DOMAIN 345 407 IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN 426 522 FIBRONECTIN TYPE-III 1.
FT DOMAIN 525 618 FIBRONECTIN TYPE-III 2.
FT DOMAIN 619 716 FIBRONECTIN TYPE-III 3.
FT DOMAIN 722 816 FIBRONECTIN TYPE-III 4.
FT DOMAIN 840 940 FIBRONECTIN TYPE-III 5.
FT DOMAIN 941 1042 FIBRONECTIN TYPE-III 6.
FT DISULFID 61 117 BY SIMILARITY.
FT DISULFID 161 212 BY SIMILARITY.
FT DISULFID 261 310 BY SIMILARITY.
FT DISULFID 352 400 BY SIMILARITY.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 318 318 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 628 628 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 168 168 M -> T (IN OESOPHAGEAL CARCINOMA).
FT VARIANT 201 201 /FTIG=VAR_003909.
FT VARIANT 1375 1375 /FTIG=VAR_003910.
FT VARIANT 138 138 P -> H (IN A COLORECTAL CARCINOMA).
FT CONFLICT 233 329 MISSING (IN REF. 3).
FT CONFLICT 421 421 MISSING (IN REF. 3).
SQ SEQUENCE 1447 AA; 158456 MW; 4A8612766ED0471F CRC64;

```

```

alignment_scores:
  Quality: 115.00      Length: 563
  Ratio: 0.475        Gaps: 26
  Percent Similarity: 42.984      Percent Identity: 20.426

alignment_block:
US-09-049-696-19 x DCC_HUMAN
Align seg 1/1 to: DCC_HUMAN from: 1 to: 1447

```

```

27  GTCCGTTTGGGGCCCTTCGACGTCAAGAACTAGAGGAGGTGTCACAAAT 76
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
189  ValValValLeuProSerGlyAlaLeuGlnIleSerArgLeu.....G1 203
77  GACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGAACATGGCC 126
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
203  nProGlyAspIleGlyIleTyrArg..... 211
127  TCATTGATGCTTTTGGGGCCCTTCATCAGCAAGATGAGCTGCTCTCAG 176
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
212  ..CysSerAlaArgAsnProAlaSerSerArgThrGlyAsnGluAlaGlu 227
177  CGCTCCATCCAGCTTGAGAGTAAGGATTA.....ACCTCCAGAACAG 220
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
228  ValArgIle...LeuSerAspProGlyLeuHisArgGlnLeuTyrPheLe 243
221  CCAGTGGATGAATGGCACAGTATGCTGGACAGCACCGTGGGAAGGACA 270
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
243  uGlnArgProSerAsnValAlaIleGlu.....GlyLysAsp 257
271  CTTTGTGTTTCTATCACCTGGACACGAGCGCTCCCAAACTCTCTCTGG 320
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
257  laValLeuGluCysValSerGlyTyrProProProSerPheThrTrp 273
321  GATCCAGTGACAGACAGAGAGGTGGTGTGTGTGTGTGTGTGTGTGTGT 370
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
274  .....LeuArgGlyGluGluValIleGlnLeuArgSerLy 285
371  AATGGCTACTCTCCAAATCCAGCATGCTAAGTTGGCAGCTTGGAAAT 420
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
285  s.....Lys 287
421  ACAGTCTGCAAGCAAGCTCAAAACCTTGACCTGACTGTACGTCGCCGT 470
    |||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

287  yrSerLeuIleGlySerAsnLeuLeuIleSerAsnValThrAsp... 302
471  GCGTCCAATGCTACCGCTGCCTCCAATTACAGTACGTTCACAAACGACAA 520
302  ..... 302
521  GGACACCAGCAAAATTCGCCAGCCCTCTGGTAGTTTATGCAAAATATCGCC 570
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
303  .AspAspSerGlyMetTyrThrCysValValThrTyrLysAsnGluAsnI 319
571  AAGGAGCCTCCCAATCTCAGGCGCAGTGTACAGCCCTGATTGAA... 617
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
319  leSerAlaSer.....AlaGluLeuThrValLeuValProPro 331
617  ..... 617
332  TrpPheLeuAsnHisProSerAsnLeuTyrAlaTyrGluSerMetAspI 348
618  .....TCAGTGAATGAAAAACAGTT...ACCTGCAACTAC 651
349  eGluPheGluCysThrValSerGlyLysProValProThrValAsnTrpM 365
652  TGGATATATGGA.....GCA 665
365  etLysAsnGlyAspValValIleProSerAspTyrPheGlnIleValGly 381
665  GGTGCTGAT.....GCTACTAAGGATGACGGTGTCTA 697
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
382  GlySerAsnLeuArgIleLeuGlyValValLysSerAspGluGlyPheTy 398
697  CTCGAAGGTATTCACACATTTATGACACGCAATGTTAGATACAGTAAAA 746
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
398  rGlnCysValAlaGluAsnGluAlaGlyAsnAlaGlnThrSerAlaGlnL 415
747  .....GTGCGGCTCTGGGAGGAGTTAAGCGCAGCC 776
415  euIleValProLysProAlaIleProSerSerSerValLeuProSerAla 431
777  AGACGGAGAGTGATACCCCGACAGAGTGGACACTGTACATACCTGGCTG 826
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
432  ProArgAspValValProValLeuValSerSerArgPheVal..... 445
827  GATTGAGAATGATCAATACAATGCAATCCACCAAGACCTGAAATTAATA 876
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
446  .....ArgLeuSerTrpArgProProAlaGluAla.....L 456
877  AGGATGATGTTCAACACACAGCAAGTGTGTTTCAGCAGA..... 914
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
456  YSGlyAsnIleGlnThrPheThrValPhePheSerArgGluGlyAspAsn 472
915  .....ACATCCTCGGAGGCTCATTT.....GT 937
473  ArgGluArgAlaLeuAsnThrThrGlnProGlySerLeuGlnLeuThrVa 489
938  GGCTTCGTGATGCCAAATGCT..... 959
489  lGlyAsnLeuLysProGluAlaMetTyrThrPheArgValValAlaTyrA 506
960  .....CCATA..... 965
505  snGluTrpGlyProGlyLysSerSerGlnProIleLysValAlaThrGln 522
965  CCGTATCTCTCCACCTGCGCAAAATCACCAGCTGAAGCGGGAATTC 1015
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
523  ProGluLeuGlnValProGlyProValGluAsnLeuGlnAlaValSerTh 539
1015  CCGGGCGCAGTCTCATTAATCTGACTTGCACAGCTCCTGGGATGATTATG 1065
539  rSerProThrSerIleLeuIleThrTrpGluProProAla.....TyrA 554
1065  ACCATGAACAGCTCACAAAGTATATCATTT.....CGAATAAGTACA 1106
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
554  laAsnGlyProValGlnGlyTyrArgLeuPheCysThrGluValSerThr 570

```

CC HYDROLASES)

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; L25809; AAC36898.1; -
DR HSSP; P07986; 1EXH.
DR InterPro; IPR001919; CBD_2.
DR InterPro; IPR001777; FN.III.
DR InterPro; IPR001524; Glyco_hydro_6.
DR Pfam; PF00553; CBD_2; 1.
DR Pfam; PF00041; Fng3; 3.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR PRINTS; PR00014; FNTYPEIII.
DR PRINTS; PR00733; GLHYDRLASE6.
DR ProDom; PD003733; Glyco_hydro_6; 2.
DR SMART; SM00060; FN3; 3.
DR PROSITE; PS00561; CBD_BACTERIAL; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
DR KEGG; Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.

FT SIGNAL 1 40
FT CHAIN 41 872 EXOGLUCANASE A.
FT DOMAIN 41 477 CATALYTIC.
FT DOMAIN 478 563 FIBRONECTIN TYPE-III 1.
FT DOMAIN 573 664 FIBRONECTIN TYPE-III 2.
FT DOMAIN 673 768 FIBRONECTIN TYPE-III 3.
FT ACT_SITE 188 188 CELLULOSE-BINDING (BY SIMILARITY).
FT ACT_SITE 410 410 PROTON DONOR (BY SIMILARITY).
FT DISULFID 140 202 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 374 428 BY SIMILARITY.
FT DISULFID 770 869 BY SIMILARITY.
FT SEQUENCE 872 AA; 89300 MW; 7883B407F995533B CRC64;

alignment_scores:
Quality: 114.50 Length: 486
Ratio: 0.513 Gaps: 21
Percent Similarity: 45.885 Percent Identity: 21.193

alignment_block:
US-09-049-696-19 x GUXA_CELFI ..
Align seg 1/1 to: GUXA_CELFI from: 1 to: 872

203 ATTAACTCTCCAGAACACCCAGTCGATGATGCGACAGTCAGTCGTGACA 252
403 ILELYSPROPGLYGLUSER...ASPGLYALASERTHRSPLEPROAS 418
253 GCACGCTGGGAAGGACACACACACACACACACACACACACACACAC 281
419 NASPGINOLYLSARGPHEASARGMETCYSPPTOTHRPHEVALSERP 435
282 ATCACTCTGGACACACACACACACACACACACACACACACACACAC 322
435 roLYSLeuAsnAsnGlnLeuThrGlyAlaThrProAsnAlaProLeuAla 451
323 TCCAGTCGACAGACAGCAGTCGCTTGTAGTCGACACACACACACAC 372
452 GYGII 465
372 TGGCCCTACTCCAAATCCAGGCATTCGTAAGTTGGCACTTGGAAATAC 422
465 ..AlaTyProValIleGlyGlyThrProValGlu..... 477
423 AGTCGACGACAGCTCACAAACCTTGACCTGACTGTGACGTCGCGTCG 472

seq_name: SwissProt_39;GUXA_CELFI
seq_documentation_block:
ID GUXA_CELFI STANDARD; PRT; 872 AA.
AC P50401;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE EXOGLUCANASE A PRECURSOR (EC 3.2.1.91) (EXOCELLULOBIHYDROLASE A)
DE (1,4-BETA-CELLULOBIHYDROLASE A) (CBP95).
GN CBHA.
OS Cellulomonas fimi.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 484;
RX MEDLINE=94344030; PubMed=8065260;
RA Meinke A., Gilkes N.R., Kwan E., Kilburn D.G., Warren R.A.J.,
RA Miller R.C. Jr.;
RT "Cellulobiohydrolase A (Cbha) from the cellulolytic bacterium
RT Cellulomonas fimi is a beta-1,4-exocellulobiohydrolase analogous to
RT Trichoderma reesei CBH II.";
RT Mol. Microbiol. 12:413-422(1994).
RL [2]
RN SEQUENCE OF 41-58.
RP MEDLINE=93209933; PubMed=8458833;
RX Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RA "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
RA D (Cend), a family A beta-1,4-glucanase.";
RT J. Bacteriol. 175:1910-1918(1993)
RL -1- FUNCTION: THIS ENZYME HYDROLYSES 1,4-BETA-D-GLUCOSIDIC LINKAGES OF
CC CELLULOSE. WEAK ACTIVITY AGAINST CARBOXYMETHYLCYLLULOSE, BACTERIAL
CC MICROCRYSTALLINE CELLULOSE AND BARLEY BETA-GLUCAN. HAS ALSO WEAK
CC ENDOGLUCANASE ACTIVITY. HYDROLYSES GLUCOSIDIC BONDS WITH INVERSION
CC OF ANOMERIC CONFIGURATION.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC LINKAGES
CC IN CELLULOSE AND CELLOTETRAOSE, RELEASING CELLOBIOSIDE FROM THE NON-
CC REDUCING ENDS OF THE CHAINS.
CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY B (FAMILY 6 OF GLYCOSYL

```
478 .....AspLeuValAlaProThrValProThrGlyLeuThrAlaGlyTh 492
473 GTCCAAATGCTACCTGCTCCAAATACAGTACGTCCAAACGAAACAAG. 521
      : : : : : : : : : : : : : : : : : : : : : : : : : :
492 rThrThrAlaThrSerValProLeuSerTrpThrAlaSerThrAspAsnV 509
      : : : : : : : : : : : : : : : : : : : : : : : : : :
522 .....GACACC 527
509 alaValThrGlyTyrAspValTyrArgGlyThrThrLeuValGlyThr 525
528 AGCAAAATCCCCACG.....CCTCGGTAGTTTA 556
      : : : : : : : : : : : : : : : : : : : : : : : : : :
526 ThrAlaAlaThrSerTyrThrValThrGlyLeuThrProAlaThrAlaTy 542
557 TGCA.....AATATTCGCAAGAGCGCTCCCAATTTCTCAGGCCCA 597
      : : : : : : : : : : : : : : : : : : : : : : : : : :
542 rSerPheThrValArgAlaLysAspAlaAlaGlyAsnValSerAlaAla 559
598 GTGTCACAGCCCTGATTGATCATCATGTAATGGAATAACA.....GTTACC 641
      : : : : : : : : : : : : : : : : : : : : : : : : : :
559 erAlaAlaAlaAlaAlaThrThrGlnSerGlyThrValThrAspThrThr 575
642 TTGGAACTACTGGATAATGAGCAGCGTGTGCTGATCTACTAAGGATGACGG 691
      : : : : : : : : : : : : : : : : : : : : : : : : : :
576 AlaProSerValProAlaGlyLeuThrAlaGlyThrThrThrThrTh 592
692 TGCTCTACTCAAGGTATTTACACTTATGACACGAATGGTAGATACAGTG 741
      : : : : : : : : : : : : : : : : : : : : : : : : : :
592 rValProLeuSerTrpThrAlaSerThrAspAlaGlyGlySerGlyV 609
742 TAAAA.....GTGCGGGCTCTGGGAGGAGTTAAC 770
      : : : : : : : : : : : : : : : : : : : : : : : : : :
609 alaAlaGlyTyrGluValLeuArgGlyThrThrValValGlyThrThrThr 625
771 GCAGCCACAGCGAGAGTGATACCCACGACAGTGGAGCAGCTGATACATACC 820
      : : : : : : : : : : : : : : : : : : : : : : : : : :
626 AlaThrSerTyrThrValThrGlyLeuThrAlaGlyThrThrTyr..... 640
821 TGGCTGGATTGAGAATGATGAATACATCAATGGAATCCACCAAGACCTGAAA 870
640 ..... 640
871 TTAATAGGATGATGTTCAACAACAAGCAAGTGTGTTTCAGCAACAATCC 920
      : : : : : : : : : : : : : : : : : : : : : : : : : :
641 .....SerPheSerValArgAlaLysAspValAla...GlyAsnThrSer 654
921 TCGGAGGCGCTCATTTGCGCTCTGATGTCCTCCCAATGCTCCCATACCTGA 970
      : : : : : : : : : : : : : : : : : : : : : : : : : :
655 AlaAlaSerAlaAlaValSerAlaThrThrGlnThrGlyThrValValAs 671
971 TCCTCTCCACCTGGCCAAATCACCGACCTGAAGGGGGAATTCACGGGG 1020
      : : : : : : : : : : : : : : : : : : : : : : : : : :
671 pThrThrAlaProSerValProThrGlyLeuThrAlaGlyThrThrThr 688
1021 GCAGTCTCATTAATCTGACTGTGACAGCTCTCTGGGATGATTATGACCAT 1070
      : : : : : : : : : : : : : : : : : : : : : : : : : :
688 hrSerSerValProLeuThrTrpThrAlaSerThrAspAsnAlaGlyGly 704
1071 GGAACAGCTCACAGTATATCATTCGAATAAGTACAAGTATTTCTGTATCT 1120
      : : : : : : : : : : : : : : : : : : : : : : : : : :
705 SerGlyValAlaGlyTyr..... 710
1121 CAGACACAAGTTCAATGAATCTCTTCAAGTGAATACTACTGCTCATTC 1170
      : : : : : : : : : : : : : : : : : : : : : : : : : :
711 ....GluValPheAsnGlyThrThrArgValAlaThrVal..... 722
1171 CAAAGGAAGCAACTCTGAGGAAGTCTTTTGTGTTTAAACCAAGAAACATT 1220
722 ..... 722
1221 ACTTTTCAAAATGGCACAGATCTTTTCATTGCTATTTCAGCGCTGTGAT.. 1268
      : : : : : : : : : : : : : : : : : : : : : : : : : :
723 .....ThrSerThrSerTyrThrValThrGlyLeuAlaAlaAspTh 736
```

seq_name: SwissProt_39:MGPC_MYCPN

seq_documentation_block:

ID MGPC_MYCPN STANDARD; PRT; 1218 AA.

AC Q50341;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE MGPC PROTEIN PRECURSOR.

GN MGPC OR MPN142 OR MP012.

CS Mycoplasma pneumoniae.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;

OC Mycoplasmataceae; Mycoplasma.

OX NCBI_TaxID=2104;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 29342 / M129;

RX MEDLINE=89211947; PubMed=2468577;

RA Inamine J.M., Loechel S., Hu P.C.;

RT "Analysis of the nucleotide sequence of the P1 operon of Mycoplasma

pneumoniae.";

RL Gene 73:175-183(1988).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 29342 / M129;

RX MEDLINE=97105885; PubMed=8948633;

RA Hummelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,

Herrmann R.;

RT "Complete sequence analysis of the genome of the bacterium Mycoplasma

pneumoniae.";

RL Nucleic Acids Res. 24:4420-4449(1996).

CC [1]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: M21519; AAA8326.1; ..

KW EMBL: AE000002; AAB95660.1; ..

KW Cell adhesion; Signal; Membrane; Complete proteome.

FT SIGNAL 1 25 POTENTIAL.

FT CHAIN 26 1218 MGPC PROTEIN.

SQ SEQUENCE 1218 AA; 130456 MW; 4DA29BCE41538311 CRC64;

alignment_scores:
Quality: 114.00 Length: 577
Ratio: 0.463 Gaps: 28
Percent Similarity: 42.634 Percent Identity: 20.277

alignment_block:
US-09-049-696-19 x MGPC_MYCPN ..

Align seg 1/1 to: MGPC_MYCPN from: 1 to: 1218

```
42  TCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAAATGACAGGAGTTTACA 91
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
380  AsnThrAlaGlnAlaLeuHisGlnMetGlyValIleValProGlyLeuG1 396
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
92  GACATATGCTTCAGATCAAGTTCCAGAAATGCGC..... 125
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
396  ulysrtpGlyGly.....AsnAsnGlyThrGlyValValAlaAs 409
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
126  .....CTCATGTGATGCTTTGGGGCCCTT 149
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
409  erArgGlnAspAlaThrSerThrAsnLeuProHisAlaAlaGlyAlaSer 425
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
150  TCATCAGGAAATCGA.....GCTGT 169
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
426  GlnThrGlyLeuGlyThrGlySerProArgGluProAlaLeuThrAlaTh 442
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
170  CTCCTACGCCCTCCATCCAGCTTGAGAGTAAGGATTAACTCCAGAAACA 219
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
442  rSerGlnArgAlaValThrValValAlaGlyProLeuArgAlaGlyAsnS 459
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
220  GCCAG..... 224
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
459  erSerGluThrAspAlaLeuProAsnValIleThrGlnLeuTyrHisThr 475
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
225  .....TGGTGAATGCGCACAGTCATCGTG..... 248
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
476  SerThrAlaGlnLeuAlaTyrLeuAsnGlyGlnIleValIleMetGlySe 492
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
249  .....GACAGCACCG 258
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
492  rAspArgValProSerLeuTyrTrpValValGlyGluAspGlnGluS 509
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
259  TGGGAAGGACACTTGTTCATACCTGGACACGACGACCTCCCAA 308
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
509  erGlyLys.....AlaThrTrpTrpAlaLysThrGlu... 519
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
309  ATCCTTCTCTGGATCCCACTGGACAGCAAGGT..... 344
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
520  ...LeuAsnTrpGlyThrAspLysGlnLysGlnPheValGluAsnGlnLe 535
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
345  .GCCTTTGTAGTCGACAAAAC..... 365
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
535  uGlyPheLysAspAspSerAsnSerAspSerLysAsnSerAsnLeuLysA 552
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
366  .....ACCAATGGCTACTCCAAATCCAGGCATT..... 398
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
552  laGlnGlyLeuThrGlnProAlaTyrIleu...IleAlaGlyLeuAspVal 567
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
399  .....GCTAAGTTGGCAGCTTGGAAATA 421
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
568  ValAlaAspHisLeuValPheAlaAlaPheLysAlaGlyAlaValGlyTy 584
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
422  CAGTCTG.....CAAGCAAGCTCAC 441
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
584  rAspMetThrThrAspSerSerAlaSerThrTyrAsnGlnAlaLeuAlaT 601
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
442  AAACCTTGACCTGACGTGTCAGCTCCGGTCCCAATGCTACCTCCCTGC 491
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
601  rpSerThrAlaGlyLeuAspSerAspGlyGlyTyrLysAlaLeuVal 617
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
492  CCAATTACAGTCACTCCAAA.....ACGACAA 520
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
618  GluAsnThrAlaGlyLeuAsnGlyProIleAsnGlyLeuPheThrLeuLe 634
      ::::::::::::::::::::: ||| ||| ||| ||| ||| |||
```

```
521  GGACACCCAGCAAAATCCCCAGCCCTCTGTAGTTATGCAAAATATTCGCC 570
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
634  uAspThrPheAlaTyrValThrProVal.....SerGlyMetLysG 648
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
571  AAGGAGCCTCCCAATTCAGGGCCAGTGTACAGCCCTCTGATTGAATCA 620
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
649  lyGlySerGlnAsnAsnGluGluValGlnThrThrTyrProValLysSer 664
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
621  GTAATGGAAAAACAGTACCTTGGAACTACTGATAATAATGGACAGGTGC 670
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
665  AspGlnLysAlaThrAlaLysIleAlaSerLeuIleAlaSerProLe 681
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
671  TGATGCTACTAAGTACAGCGGTCTCTACTCAAGTATTTCACAACTATG 720
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
681  uAsnSerTyrGlyAspAspGlyVal.....ThrValPheA 693
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
721  ACAGGAATGCT...AGATACAGTGTAAAGTGGCGCTCTGGGAGGAGTT 767
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
693  spAlaLeuGlyLeuAsnPheAsnPheLys.....Leu 703
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
768  AAGCAGCCAGAGGAGTATACCCAGCAGAGTGGAGCAGCTGTACAT 817
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
704  AsnGluGluArg.....LeuProSerArgThrAspGlnLeuLeuVa 717
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
818  ACCTGGCTGGATTGAGATGATGAATACAATGGAATCCACCAAGACCTG 867
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
717  lTyrGlyIleValAsnGluSerGluLeuLysSerAlaArgGluAsnAlaG 734
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
868  AAATTAATAAGGATGTTCAACACAGCAAGTGTGTTTTCAGCAGAAACA 917
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
734  lnSerThrSerAspAsnSerAsnThrLysValLysTyrThrAsnThr 750
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
918  TCCTCGGAGGCTCATTTGTGGCTCTCTGATGTCCTCCAAATGCTCCATCC 967
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
751  AlaSer.....HistyLeuProValPr 758
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
963  TGATCTC.....TTCCACCTGGC..... 986
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
758  oTyrTyrTyrSerAlaAsnPheProGluAlaGlyAsnArgArgAlaG 775
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
987  .....CAAATCACCAGCTGAAGCGGAAATTCACGGG 1019
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
775  luGlnArgAsnGlyValLysIleSerThrLeuGluSerGlnAlaThrAsp 791
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
1020  GGC.....AGTCTCATTAATCTG...ACTTGGACAGCTCTCTGGGGA 1057
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
792  GlyPheAlaAsnSerLeuLeuAsnPheGlyThrGlyLeuLysAlaGlyVa 808
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
1058  TGAT.....TATGACCATGGAACACAGCTCACAAAGTATATCA 1092
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
808  lAspProAlaProValAlaArgGlyHisLysProAsnTyrSerAlaValL 825
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
1093  TTCGAATAAGTACAGTATCTTGTATCTCAGA.....GAC 1127
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
825  euLeuValArgGlyGlyValValArgLeuAsnPheAsnProAspThrAsp 841
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
1123  RAGTTCAATGAATCTCTTCAAGTGAATACTACTGCTCTCTCTCCCAAGGA 1177
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
842  LysLeuLeuAspSerThrAspLysAsnSer..... 851
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
1173  AGCCAACTCTGAGGAAGTCTTTTGTAAACACAGAAACATTACTTTTG 1227
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
853  .....GluProIleSerPhe. 856
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
1223  AAAATGGCACAGATCTTTTCATTGCTATTACAGCTGTTGATAAGTCGAT 1277
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
857  ..SerTyrThrProPheGlySerAlaGluSerAlaValAspLeuThrThr 872
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
1278  CTGAATCAGAAATATCCAACTGACAGGATGATCTTTGTTTATTCCTCC 1327
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
873  LeuLys...AspValThrTyrIleAlaGluSerGlyLeuTyrPheTyrTh 888
      ||| ||| ::::::::::::::::::::: SerGlyMetLysG 648
```



```

731 TAGATACAGTGAAGTGGGGCTGTGGAGGAGTTAACGACCCAGAC 780
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
294 yAsnLysThrIleGluAlaGlnAlaGluLeuThrVal..... 306
781 GGAGAGTATACCCAG.....CAGAGTGAGCACTGTAC..... 815
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
307 ..GlnValProGluLeuLysGlnProAlaAsnIleTyrAlaArg 322
816 .....ATACCTGGC..... 824
323 GluSerMetAspIleValPheGluCysGluValThrGlyLysProAlaPr 339
825 .....TGGATTGAGATGATGAATA..... 845
339 oThrValLysTrpValLysAsnGlyAspValValIleProSerAspTyrP 356
846 .....CAATGGAATCCACCAAGACCTCGAAATTAATAGGAT 881
356 heLysIleValLysGluHisAsnLeuGlnValLeuGlyLeuValLysSer 372
882 GATGTTCAACACCAAGTGTGTTACGACAGACATCTCGGAGGCTC 931
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
373 AspGluGlyPheTyrGln...CysIleAlaGluAsnAspValGlyAsnAl 388
932 ATTGTGGCTTCTGAT.....GTCCCAATGCTCCCATCTGATC 972
: : : : : : : : : : : : : : : : : : : : : : : :
388 aGlnAlaGlyAlaGlnLeuIleLeuGluHisAlaProAlaThrThrG 405
973 TCTTCCCACTGGCAATACCCAGCTTGAAGCGGAATTCACGGGGC 1022
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
405 lProLeuProSerAlaProArgAspValAlaSerLeuValSerThr 421
1023 AGTCTCATTAATCTGACTTGGACAGCTCTGGGATGATATGACCATG 1072
: : : : : : : : : : : : : : : : : : : : : : : :
422 ArgPheIleLysLeuThrTrpArgThrProAlaSerAsp...ProHisG 437
1073 ACAGCTCACAGTATATCATTCGTAATAGTACAGTATCTTGATCTCA 1122
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
437 yAspAsnLeuThrTyrSerValPheTyrThrLysGluGlyValAlaAa+gg 454
1123 GAGACAAGTTCATGAATCT.....CTTCAAGTGAATACTACT 1160
: : : : : : : : : : : : : : : : : : : : : : : :
454 lArgValGluAsnThrSerGlnProGlyGluMetGlnValThrIleGln 470
1161 GCCTCATCCCAAGGAAGCAACTCTGAGGAAGTCTTTTGTAAACC 1210
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
471 AsnLeuMetPro.....AlaThrValTyrIlePheLys... 481
1211 AGAAACATTACTTTTGAANAATGGCACAGATCTTTTCATTGCTATT 1260
482 .....ValMeCa 484
1261 CTGTGTGATAAGTCGATCTGAATCAGAATAATCCAACTGCACAGTA 1310
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
484 lArgAsnLysHisGly...SerGlyGluSerSerAlaProLeuArgVal 499
1311 TCTTTGTTTATCTCTCCAGACCTCGCCAGACACCTAGTCTCATCA 1360
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
500 .....GluThrGlnProGluValGlnLeuPro..... 508
1361 AACGTCTGCTCTGTCTCTAATATTCATATCAACAGCACCATTCCTGG 1410
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
509 .....GlyProAlaProAsnIleArgAlaTyrAlaThrSerPro.... 521
1411 TTCATTTTAAATAATCTGGAAG.....TGGATAGAGAA 1448
|::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
522 ..ThrSerIleThrValThrTrpGluThrProLeuSerGlyAsnGlyGlu 537
1449 CTGCAG 1454
: : : : : : : : : : : : : : : : : : : : : : : :
538 IleGln 539

```

seq_name: SwissProt_39:EAE_ECO57

```

seq_documentation_block:
ID EAE_ECO57 STANDARD; PRT; 934 AA.
AC P43261; 085627;
DT 01-NOV-1995 (Rel. 32, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTIMIN (ATTACHING AND EFFACING PROTEIN) (EAE PROTEIN) (GAMMA-
DE INTIMIN).
GN EAE OR EAPA OR Z5110 OR ECS4559 OR L0025.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933;
RX MEDLINE=92204008; PubMed=1552854;
RA Yu J., Kaper J.B.;
RT "Cloning and characterization of the eae gene of enterohaemorrhagic
RT Escherichia coli O157:H7.";
RL Mol. Microbiol. 6:411-417(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / CL-8;
RX MEDLINE=92249750; PubMed=1577255;
RA Beekakhee G., Louie M., de Azavedo J., Brunton J.;
RT "Cloning and nucleotide sequence of the eae gene homologue from
RT enterohaemorrhagic Escherichia coli serotype O157:H7.";
RL FEMS Microbiol. Lett. 70:63-68(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 700927;
RX MEDLINE=98339885; PubMed=9673266;
RA Peina N.T., Mayhew G.F., Posfal G., Elliott S., Donnenberg M.S.,
RA Kaper J.B., Blattner F.R.;
RT "Molecular evolution of a pathogenicity island from enterohaemorrhagic
RT Escherichia coli O157:H7.";
RL Infect. Immun. 66:3810-3817(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / DEC 3a, AND O157:H- / DEC 3f;
RX MEDLINE=99261633; PubMed=10331248;
RA McGraw E.A., Li J., Selander R.K., Whittam T.S.;
RT "Molecular evolution and mosaic structure of alpha, beta, and gamma
RT intimins of pathogenic Escherichia coli.";
RL Mol. Biol. Evol. 16:12-22(1999).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Gröbbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kohara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [7]
RP SEQUENCE OF 1-37 FROM N.A.
RC STRAIN=O157:H7 / HAI;

```

[illegible]

708 sSerGlnThrGlnAlaThrThrGlyAsnAspGlyArgAlaThrIleThrL 725
740 TGTAAGAAGTCGGGCTCTGGGAGGAGTTAAACGCAGCCAGACGAGACTGA 789
725 eu.ThrSerSerAlaGlyLysAlaThrValSerAlaThrValSerAs 741
790 TACCCAGCAGAGTGGGCACTGTACATACCTGGCTGGATTGAG..... 833
741 pGlyAlaGluValLysAlaThrGluValThrPhePheAsp.GluLeuLys 757
834AATGATGAATACAAATCGAATCCACCAGACCTGAAATTAA 874
753 IleAspAsnLysValAspIleIleGlyAsnValArgGlyGluLeuPr 774
875 TAAGGATGATCTTCCACACAGCAAGTGTGTTTCAGCAGCAACATCTCCFCGG 924
774 oAsnIleTriLeuGlnThrGlyGln.....PheLysLeuLysAlaSerG 789
925 GAGGCTCATTTGTGGCTTCT...GATGCCAAATGCTCCCATACCTGAT 971
789 lyGlyAspGlyThrTyrSerTriTyrSerGluAsnThrSerIleAlaThr 805
972 CTCCTCCACCTCGGCCAAATCACC.....GACCT 1000
805 ValAspAlaSerGlyLysValThrLeuAsnGlyLysGlySerValValII 822
1001 GAAGCGGAAATTACGGGGGCAGCTCTCATTAATCTGACTTGCACAGCTC 1050

```

MEDLINE=20196006; PubMed=10731132;
Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yeates M.D., Zhang Q., Chen L.X.,
Wang K.H., Rogers Y.-H.C., Blazie R.G., Champe M., Pfeiffer B.D.,
Bridgton R.C., Rogers J.E., McEwen G.S., Olsen V.J., McKnight D.L.,
Mills A.P., Moore T., Orkin S.C., Poulos B., Rothberg L.M.,
Spradling A.C., Staden R., Strydom E., Thayer M.S., White O.,
Yee N.K., Zeng H., Zhao Q., Zhong W., Zhou X., Zhu S., Smith H.O.,
Fink D.L., Davis R.P., Myers E.W., Rubin G.M., Venter J.C.; "The
genome sequence of Drosophila melanogaster." Science
287:2185-2195(2000). [5]
SEQUENCE OF 1182-1302 FROM N.A., AND CHARACTERIZATION.
Tissue=Embryo; MEDLINE=90262720; PubMed=1693086;
Hortsch M., Bieber A.J., Patel N.H., Goodman C.S.;
"Differential splicing generates a nervous system-specific form of
Drosophila neuroglian." Neuron 4:697-709(1990). [6]
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 610-814.
MEDLINE=94213741; PubMed=7512815;
Huber A.H., Wang Y.-M.E., Bieber A.J., Bjorkman P.J.;
"Crystal structure of tandem type III fibronectin domains from
Drosophila neuroglian at 2.0 A." Neuron 12:717-731(1994).
FUNCTION: THE LONG ISOFORM MAY PLAY A ROLE IN NEURAL AND GLIAL
CELL ADHESION IN THE DEVELOPING EMBRYO. THE SHORT ISOFORM MAY BE A
MORE GENERAL CELL ADHESION MOLECULE INVOLVED IN OTHER TISSUES AND
IMAGINAL DISK MORPHOGENESIS. VITAL FOR EMBRYONIC DEVELOPMENT.
SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
TISSUE SPECIFICITY: LONG ISOFORM IS RESTRICTED TO SURFACE OF
NEURONS AND GLIA IN THE DEVELOPING NERVOUS SYSTEM AND THE SHORT
ISOFORM TO OTHER NONNEURAL TISSUES.
SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sb.ch/announcement/)

```


Tue Apr 2 09:39:54 2002

us-09-049-696-19.rsp

```
776 CAGACGGAGAGTGATA.....CCCCAG. 797
    |||:|
512 pGlySerLeuValValLysGluHisThrArgIleThrGlnGluProGlnA 529
798 .....CAGAGTGGAGCA..... 809
529 snTyrGluValAlaAlaGlyGlnSerAlaThrPheArgCysAsnGluAla 545
810 .....CTGTACATACCTGGGTGATTGAGATGATGA 841
546 HisAspAspThrLeuGluIleAspTrpIleAspGlyGlnse 562
842 ANTACATGGATCCACCAAGCTGAATTAATAAGATGAT..... 884
562 rIleAspPheGluAlaGlnProArgPheValLysThrAsnAspAsnSerL 579
885 .....GTTCAACACAAAGCTGTGTTC 908
579 euThrIleAlaLysThrMetGluLeuAspSerGlyGluTyrThrCysVal 595
909 ACAGAGACA.....TCCTCGGAGGCTCATTTGGCTTC 943
596 AlaArgThrArgLeuAspGluAlaThrAlaArgAlaAsnLeuIleValGI 612
944 TGATGTCCCAATGCTCCATACCTGATCTCTCCACCTGGCCAAATCA 993
612 nAspValProAsnAlaPro.....LysLeuT 621
994 CCGACCTG.....AAGCGGAAATTCACGGGGCAGTCTC 1028
621 hrGlyIleThrCysGlnAlaAspLysAlaGluIleHis..... 633
1029 ATTAATCTCAGCTGGACAGCTCCTGGGATGATTATGACCATGGAACAGC 1078
634 .....TrpGluGlnGlyAspAsn..... 640
1079 TCACAAGTATATCATCGAATAAGTACAGATATCTTGTATCTCAGAGACA 1128
641 .....ArgSerProIleLeuHisTyrThrIleG 650
1129 AGTTCAATGAATCTCTT...CAAGTCAATACTACTGCTCTCATCCCAAAG 1175
650 InPheAsnThrSerPheThrProAlaSerTrpAspAlaAlaTyrGluLys 566
1176 GAAGCCAACTCTGAGGAAGCTTTTGTTTAAACCAGAA.....AA 1216
667 ValProAsnThrAspSerSerPheValValGlnMetSerProTrpAlaAs 683
1217 CATTACTTTGAAATGGCACAGATCTTTTCATTGCTATTCAGGCTGTG 1266
683 nTyrThrPheArg.....ValIleAlaPheA 692
1267 ATAAGGTCGATCGAATCAGAAATATCCAACATTCACGAGTATCTTTG 1316
692 snLysIleGlyAlaSer..... 697
1317 TTTATTCTCCACAG.....ACTCCGCCACAGAC 1345
698 .....ProProSerAlaHisSerAspSerCysThrGlnProAspVa 712
1346 ACCT.....AGTCTGATGAA..... 1361
712 lProPheLysAsnProAspAsnValValGlyGlnGlyThrGluProAsnA 729
1362 .....ACGCTGCTCCTGCTTAATATTCATATCAACAGACCC 1400
729 snLeuValIleSerTrpThrProMetProGluIleGluHisAsnAla... 744
1401 ATTCTGCGCATTCACATTTTAAAAATTTATGTGGAAG 1436
745 ...ProAsnPheHisTyrTyr...ValSerTrpLys 754
```

OM of: US-09-049-696-19 to: PIR_68.* out_format : pfs

Date: Mar 30, 2002 2:31 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=framer_nsp.model -DEV=xlp
-Q/cgn2.1/uspto_spool/us09049696/runat_28032002_145238_2031/app_query.fasta_1.12579
-DB=PIR_68 -QFMT=fastan -SUFFIX=pr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USR=US09049696_@CGN1_1_367 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-049-696-19
Query length: 1683
Database: PIR_68.*
Database sequences: 219241
Database length: 7617452
Search time (sec): 470.790000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
pir2:TI4160	+	135.50	0.0311	1651	transmembrane receptor protein
pir2:B85547	+	134.50	0.0413	5188	probable RTX family exoprotein
pir2:150620	+	134.00	0.0350	1173	procr2 - chicken (fragment)
pir2:T21810	+	132.50	0.0414	683	hypothetical protein F35G12.4b
pir2:T21808	+	131.00	0.0539	686	hypothetical protein F35G12.4a
pir2:C81675	+	130.00	0.0747	1672	polymorphic membrane protein B
pir2:T30805	+	128.00	0.1052	1612	gut1 protein - mouse
pir1:D54689	+	127.50	0.1158	1691	protein-tyrosine-phosphatase
pir2:C54689	+	127.50	0.1180	1894	protein-tyrosine-phosphatase
pir2:A38096	+	124.00	0.1487	4391	perlecan precursor - human
pir2:A56178	+	124.00	0.2178	1912	protein-tyrosine-phosphatase
pir1:I38344	+	121.00	0.1715	26926	titin, cardiac muscle [invalid
pir2:S84023	+	120.50	0.3097	417	probable transcription factor M
pir2:B35621	+	119.50	0.3844	532	spore germination protein 270-1
pir2:T02759	+	119.50	0.4218	918	hypothetical protein - Acinetob
pir2:S74325	+	118.50	0.4594	250	hypothetical protein sir0431
pir2:T32828	+	117.50	0.6725	1825	hypothetical protein H19M22.1
pir2:CA1859	+	117.50	0.6740	1849	IgA-specific metalloendopeptid
pir2:F69009	+	115.50	0.9195	1474	probable membrane protein MTH
pir2:A34100	+	115.00	1.00	1447	tumor suppressor protein DCC p
pir2:S49541	+	114.50	1.00	872	cellulase - Cellulomonas fimi
pir2:S20901	+	114.00	1.16	1218	hypothetical pl operon 130K pr
pir2:T24643	+	113.50	1.19	6805	titin - rabbit (fragment)
pir2:E85822	+	113.00	1.33	1007	hypothetical protein T07C12.8
pir2:B81281	+	112.50	1.57	2660	probable invasin 23135 [import
pir2:C86045	+	112.50	1.46	1041	probable secreted serine prote
pir2:A29355	+	110.50	1.71	934	intimin adherence protein [imp
pir2:D81675	+	110.50	1.72	1020	fibronectin - chicken (fragme
pir1:FLFC	+	110.00	1.29	1460	polymorphic membrane protein B
pir2:T23979	+	110.00	1.38	498	flagellin - Escherichia coli (S
pir2:T23986	+	110.00	1.26	1941	hypothetical protein R06F6.8a
pir1:A32579	+	109.50	1.26	252	hypothetical protein R06F6.8b
pir2:T34434	+	109.50	1.29	1239	neuroglian - fruit fly (Drosop
pir2:C81265	+	109.00	1.24	281	hypothetical protein K06A9.1a
pir2:DB1411	+	109.00	1.31	928	probable lipoprotein Cj1678 [im
pir2:PN0568	+	109.00	1.31	978	probable lipoprotein Cj0629 [im
pir2:T31102	+	109.00	1.28	1323	connectin 3B - chicken (fragme
pir2:I38346	+	109.00	1.17	4152	filamentous hemagglutinin 1 -
pir2:T34369	+	108.50	3.81	7962	elastic titin - human (fragmen
pir2:B75569	+	108.00	124.93	322	hypothetical protein Tl912.1
	+	108.00	137.88	399	conserved hypothetical protein

pir2:T25168 + 108.00 127.38 3.32 1262 ! hypothetical protein T23F1.
pir2:S59077 + 107.00 127.35 3.85 1090 ! cellulose 1,4-beta-cellobio
seq_name: pir2:TI4160
seq_documentation_block:
transmembrane receptor protein Robol - rat
C:Species: Rattus norvegicus (Norway rat)
C:date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14160
Cell 92; 205-215, 1998
R:Kidd, T.; Brose, K.; Mitchell, K.J.; Fetter, R.D.; Tessier-Lavigne, M.; Goodman, C.
A:Title: Roundabout controls axon crossing of the CNS midline and defines a novel sub
A:Reference number: Z17897; MUID:98117249
A:Accession: T14160
A:Status: preliminary;
A:Molecule type: mRNA
A:Residues: 1-1651 <KID>
A:CROSS-references: EMBL:AF041082; NID:g2811215; PID:g2811216; PIDN:AAC39960.1
C:Function:
A:Description: appears to function as the gatekeeper controlling midline crossing
C:Keywords: transmembrane protein

alignment_scores:
Quality: 135.00 Length: 457
Ratio: 0.619 Gaps: 23
Percent Similarity: 46.681 Percent Identity: 21.413

alignment_block:

US-09-049-696-19 x T14160 ..
Align seg 1/1 to: T14160 from: 1 to: 1651
238 ATGATGCGACAGTGTGTCGACGACGACCGTGGAAAGGACACCTTGT 277
:::|||||:::|||||:::|||||:::|||||:::|||||
462 ValasnGlnThrValalaValaspGlyThrLeu.....ThrLeuSe 475
278 TCTTATACCTGGACAGCGACGCTCCCAATCCTCTCTGGGATCCCA 327
475 rCysValala...ThrGlySerProValProThrIleLeuTrp..... 488
328 CTGGACAGCAAGGTCGCTTGTAGTGGACAAAACACCAAAATGGCC 377
:::|||||:::|||||:::|||||:::|||||:::|||||
489ArgGlyAspGlyValLeuValSerThrGlnAspSerArgIleLys 503
378 TACTCCAAATCCCGAGCATT.....GCTAAGGTTGGC... 410
504 GlnLeuGlu...SerGlyValLeuGlnIleArgTyrAlaLysLeuGlyAs 519
411 .ACTTGGAAATACAGTCTGCAAGCAAGCTCACAACCTTGACCCCTGACTG 459
:::|||||:::|||||:::|||||:::|||||:::|||||
519 pThrGlyArgTyrThrCysThrAlaSerThrProSerGlyGluAlaThr 536
460 TCACGTCC..... 467
536 rpSerAlaTyrIleGluValGlnGluPheGlyValProValGlnProPro 552
468 CGTGGCTCCAATGTACCTGCTCCTCAATTACAGTGACTTCCAAACGAA 517
|||:::|||||:::|||||:::|||||:::|||||
553 ArgProThrAspProAsnLeuIleProSerAlaProSerLysProGluVa 569
518 CAAGGACACACGAAATTCACGACCCCTCTGTAGTTTATGCAATATTC 567
|||:::|||||:::|||||:::|||||:::|||||
569 lThrAspValSerLysAsnThrValThrLeuLeuTrpGlnProAsnLeuA 586
568 CCCAAGGACCTCCCA.....ATTCTCAGGGCC..... 596
586 snSerGlyAlaThrProThrSerTyrIleIleGluAlaPheSerHisAla 602
597ACTGTACACGACCTGATTCAATCAATGAGTGAATAACAGT 637
|||:::|||||:::|||||:::|||||:::|||||
603 SerGlySerSerTrpGlnThrValAlaGluAsnValLysThrGluThrPh 619

```
638 TACCTTGGAA.....CTACTGGATAATG 660
        :::::
619 eAlaileLysGlyLeuLysProAsnAlaileTyrLeuPheLeuValArgA 636
        :::::
661 GACAGGTGCTGATCTACTAAGATGACGGTCTCTACTCAAGGTATTTC 710
        :::::
636 lAlaAsnAlaTyrGlyIleSerAspProSerGlnIleSerAspProVal 652
        :::::
711 ACAACTTATGAC.....ACGAATGGTAGATACAGTGTAAAGT 748
        :::::
653 LysThrGlnAspValProProThrThrGlnGlyValAspHisLysGlnVa 669
        :::::
749 G...CGGGCTCGGAGGAGTT..... 767
        :::::
669 lGlnArgGluLeuGlyAsnValValLeuHisLeuHisAsnProThrIleL 686
        :::::
768 .....NACGGACGCAGAGGAGTAGATACCCAGCAGAGT... 803
        :::::
686 euSerSerSerValGluValHisThrThrValAspGlnGlnSerGln 702
        :::::
804 .....GGACACTGTATACATCCCTGGCTGGATTGAGATGA 838
        :::::
703 TyrIleGlnGlyTyrLysIleLeuTyrArgProSerGlyAlaSerHisG 719
        :::::
839 TGAATACATGG.....ATCCA..... 857
        :::::
719 yGluSerGluTrpLeuValPheGluValArgThrProThrLysAsnSer 736
        :::::
858 .....CCAAAGA 863
        :::::
736 alValIleProAspLeuArgLysGlyValAsnTyrGluIleLysAlaArg 752
        :::::
864 CCTGAAATTAATAGGATGATGTTCAACACAAGCAAGTGTGTTTCAGCAG 913
        :::::
753 ProPhePheAsnGluPheGlnGlyAlaAspSerGluIleLysPheAlaY 769
        :::::
914 ACATCTCTCGGGAGGCTATTTGGCTCTCTGATGTGCCAAATGCTCCCA 963
        :::::
769 sThr.....LeuGluGluArgProSerAla.... 777
        :::::
964 TACCTGATCTCTCCACCTGGCAATATCACCGACCTGAAGCGGAAATT 1013
        :::::
778 .....ProProArgSerValThrValSerLysAsnAspGly 789
        :::::
1014 CACGGGGCAGTCTCAATTAATCTGACTGTGACAGCTCTGGGATGATTA 1063
        :::::
790 AsnGlyThrAlaIleLeu...ValThrTrpGlnProProGluAspTh 805
        :::::
1064 TGACCATGGAACAGCTCACAAATATATCATTCGANTAAAGTACAAGTATTC 1113
        :::::
805 rGlnAsnGlyMetValGlnGluTyrLysValTrpCysLeuGlyAsn.... 820
        :::::
1114 TTGATCTCAGACAGAGTTCAATGAATCTCTCAAGTGAATACTACTGTCT 1163
        :::::
821 ..GluThrArgTyrHisIleAsnLysThrValAspGlySerThrPheSer 836
        :::::
1164 .....CTCATCCCAAGGAAGCAACTCTCAGGAAGTC... 1196
        :::::
837 ValValIleProPheLeuValProGlyIleArgTyrSerValGluValAl 853
        :::::
1197 .....TTTTTGTTTAAACCAAGNAACATTA 1221
        :::::
853 aAlaSerThrGlyAlaGlyProGlyValLysSerGluProGlnPheIleG 870
        :::::
1222 CTTTGAATAATGGCACAGCTTTTCATTGCTATTACGGCTGTGTATAAG 1271
        :::::
870 lnLeuAspSerHisGlyAsn.....ProValSerProGluAspGln 883
        :::::
1272 GTCATCTGAATACAGAAATATCAACATTCACAGTACTCTTTGTTTAT 1321
        :::::
884 ValSerLeuAlaGlnGlnIleSerAspValValLysGlnProAlaPheIl 900
        :::::
```

seq_name: pir2:B85547

seq_documentation_block:

probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: B85547
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85547
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5188 <STO>
A:Cross-references: GH:AH005174; NID:gl2513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0615

alignment_scores:

Quality: 134.50 Length: 361
Ratio: 0.846 Gaps: 19
Percent Similarity: 44.044 Percent Identity: 23.546

alignment_block:

US-09-049-696-19 x B85547 ..

Align seg 1/1 to: B85547 from: 1 to: 5188

```
12 GCATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAACTAGA 61
        :::::
4144 AlaLeuIleAspThrVal.....AsnProAlaGlnValLeuThrIleAl 4158
        :::::
62 GGAGCTGTCCAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAG 111
        :::::
4158 aSerIleSerThrAspThrGly.....SerSerAlaThrAspPheI 4172
        :::::
112 TTCAGAACAAATGCCCTCATTCATGCTTTTGGGCCCTTCATCAGGA... 158
        :::::
4172 leThrSerAspThrMetLeuThrLeuThrGlySerLeuGlyAlaGlyLeu 4188
        :::::
159 ...AATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGA... 203
        :::::
4189 AlaSerGlyGluValAlaGln.....IleSerLeuAspSerGlyAlaTh 4203
        :::::
204 .....TTAACCTCCAGAACAGCCAGTGGATG..... 230
        :::::
4203 rTrpThrThrLeuThrThrAsnGlyThrGlnTrpThrThrThrAspSerA 4220
        :::::
230 ..... 230
        :::::
4220 rgThrLeuThrAspGlySerTyrValTyrGlnValArgValLeuAspLeu 4236
        :::::
231 .....AATGGCACAGTATCGTGGACAGCACCGTGGAAAGGACACTTT 274
        :::::
4237 AlaGlyAsnThrGlyProValValSerLysThrValValValAspThr.. 4252
        :::::
275 GTTCTTTATCACCTGGACAAACGAGCTCCCAATCTCTCTCTGGGATC 324
        :::::
4253 .....IleAsnProThrAlaThrProThrIleValSerTyrThrAspA 4267
        :::::
325 CCAGTCGACAGAACCAAGGTGGCTTT.....GTACTGGACAAA 362
        :::::
4267 spValGlyGlnArgGlnGlyThrLeuSerSerLeuGlnAlaThrAspAsp 4283
        :::::
363 AACACC.....AA 370
        :::::
```

```

:::||||
4284 thrThrProLeuLeuAsnGlyValLeuSerAlaProLeuAlaSerGlyG1 4300
371 AATGGCTACCTCCAAATCCAGGC.....ATTGCTAAGG 405
:::|||||
4300 uValValTyrLeuTyrArgAsnGlyLeuLeuGlyAlaValThrMetV 4317
406 TTGGC.....ACTTGGAAATACAGT.....CTCAAGCAAGCTCA 440
||||| ::||| ||||| ||| :::::
4317 alGlyAlaLeuAsnTrpThrTyrSerAspSerGlyLeuValSerGlyAla 4333
441 CAACCTTGACCTGACCTGACGTGACGTCGCTCAATGCTACCTGCC 490
||| ::::: ||| ::::: ||| ::::: |||
4334 TyrThrTyrSerAlaArgValValAspLeuAlaGlyAsnIleThrSerSe 4350
491 T.....CCAAATTACA..... 500
4350 rSerAspPheValLeuThrValAspThrSerIleProThrThrLeuAlaG 4367
||| |||
501 ..GTGACTTCCAAAGCAACAGGACACACGCAATTCGCCAGCCTCTG 548
:::||||| ::||| |||||
4367 InIleThrSerGlnThrThrArgAspThr..... 4376
549 GTAGTTTATGCAAAATATTCGCCAAGGAGCCTCCCAATTCAGGGCCAG 598
:::||||| ::|||
4377 .....ThrProIleIleSerGlyVa 4383
599 TGTACAGCCCTGATTGAATCA.....GTGAATG 627
:::||||| ::|||
4383 I1leThrAlaAlaLeuAlaSerGlyGlnTyrValGluValValIleAsnG 4400
628 GAAACACAGTTACCTTGGAACTACTGGTAATGGAGCAGCTGCTGCT 677
||||| |||
4400 lYlYsThrThrThrSerGlu.....ProGlyGlyAlaVal 4411
678 ACTAAGGATGACGGTGTCTACTCAAGGTATTTCACAACCTTATGACACAA 727
::: ||| ::|||
4412 ValValAspProAlaHisAsnThrTrpTyrValGlnLeuProAspThrAs 4428
728 TGGT.....AGATACAGTGTAAAG 747
::: |||
4428 pAlaLeuThrValSerAlaThrAlaTyrThrValThrAlaGlnValYsS 4445
748 TCCGGGCTCGGAGGAGTTAAGCA....GCCAGAGGAGAGTGATACCC 794
||| ::||| |||||
4445 erSerAlaGlyAsnGlyAsnAsnAlaAsnIleSerAsnGlyThrValThr 4461
795 CAGCAGAGTGGAGCACTGTACATACCTGGCTGG 827
::: ||| ||| |||
4462 ValAsnAlaAlaIleAspTyrThrProThrTrp 4472

```

seq_name: pir2:I50620

seq_documentation_block:

procr2 - chicken (fragment)

C:Species: Gallus gallus (chicken)

C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: I50620

R:Schutz, B.; Niessing, J.

Gene 148, 227-236, 1994

A:Title: Cloning and structure of a chicken zinc finger cDNA: restricted expression in d

A:Reference number: I50620; MUID:95047430

A:Accession: I50620

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1173 <SCH>

A:Cross-references: EMBL:X56805; NID:g577018; PIDN:CAA40140.1; PID:g577019

C:Genetics:

A:Gene: ckr2

alignment_scores:

Quality: 134.00 Length: 535

Ratio: 0.565 Gaps: 21
Percent Similarity: 44.299 Percent identity: 19.626

alignment_block:

US-09-049-696-19 x I50620

Align seg 1/1 to: I50620 from: 1 to: 1173

```

159 AATGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGGA..... 203
::: ||| ::||| ||||| ||| ::::: |||||
519 AspGlyProValLeuGlnAlaAlaLeuSerLeuGluValThrAlaProAs 535
44 .....TTAACCCCTCCAGACAGCAGTGGATGAATGCCACAG 240
||| ||| ||||| ::||| :::::
535 pAlaHisThrPheLeuLeuGlnThrProGlnGlyLeuGln.....L 550
241 TGATCGTGGACAGCACCGTGGGAMAGGACACTTGTCTTCTATCACCTGG 290
::: ||| ::||| ::||| ::||| ::||| ::|||
550 euIleProSerProProAlaProGlnLysLeuIleLeuLeuProThr 566
291 ACAAGCAGCCTCC..... 307
::: ||||| |||
567 AlaProGlnProProLysHisGlnGlyGluSerProThrProGlyG1 583
308 AATCCTTCTCTGGGATCCAGTGA.....CAGA 336
||| |||||
583 nserLeuLeuLeuValProSerThrGlyThrLeuProThrLeuArgL 600
337 AGCAAGTGGCTTTGTAGTGGACAAAACACCAAAATGGCTACCTCAA 386
||| ::||| ::||| ::||| ::|||
600 euGlnAlaValThrAlaValProGlnGlyThrGlyThrGlyIleLeuVal 616
387 ATCCAGGAGT..... 398
617 LeuGlnGlyLeuProGluGlnProLeuHisProAlaGlyIleProHisG1 633
399 ....GCTAAGTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAA 444
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
633 yGlnAlaAlaValGluGlyAlaAlaValArgLeuGlnAlaAsnGluVal 650
445 CC...TTGACCTGACTCTCAGTCCCGTGGCTCCAATGCTACCTGCT 491
|| ::||| ::||| ::||| ::||| ::|||
650 hrAsnValGlnLeuGlnAlaLeuProGlnProThrAspValThrSerIle 666
492 CCAATTACAGTGACTTCCAAAACGAAGACACCAAGACCAAAATTCGCCAG 541
||| ::||| ::||| ::||| ::|||
667 ProLeuGlnAlaThrAspMetThrAsnValGlnLeuGlnAlaLeuProG1 683
542 CCCTCTGGTGTATTATGCAAAATATTCGCCAAGGAGCCTCCCAATCTCA 591
::: ||| ::||| ::||| ::||| ::|||
683 nProThrAspVal...ThrAsnIleGlnLeuGlnAlaThrGluValThrA 699
592 GGGCAGTGTACAGCCTGATTGAATCAGTGAATGGAACAAACAGTTACC 641
::: ||| ::||| ::||| ::||| ::|||
699 snValGlnLeuGlnAlaLeuLeuGlnProThrAspValThrAsnIleGln 715
642 TTGGAACCTACTGGATAATGGAGCAGGTGCTGATGCTACTAAGGATGACGG 691
::: ||| ::|||
716 MetGlnAlaThrGluValAla..... 722
692 TGTCTACTCAAGGTATTTCAACAATTCACAGCAATTCACAGTAATGTA 741
::: |||
723 ..... 724
742 TAAAGTGGGCTCTG..... 758
::: ||| ::|||
744 alGlnIleGlnAlaLeuSerGlnProProAspValThrAsnIleGluLeu 740
759 .....GGAGGAGTTAAGCAGCCAGAGGAGTATACCCAGCAGCAG 802
::: ||| ::|||
741 GlnAlaAlaGluValThrAsnAlaHisLeuGlnValLeuProGlnProSe 757
803 TGGAGCACTGTACATA.....CCTGGCTGGATTGAGAAATGATGAAA 843

```



```
600 GTCACAGCCCTGATTGAATCACTAGTGAATGGAAGAAACAGTTACTTGGAACT 649
      ::::::::::: ::::|::|::| ::::|::|::| ::::|::|::|
374 .....ValLysAspGlnLysGlyGlnGlnSerThrProGluLeu 386
      ::::::::::: ::::|::|::| ::::|::|::| ::::|::|::|
650 ACTGATAATGAGCAGCGTGTGATGCTTACTTAAGGATCAGCGTGCTACT 699
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
386 uvalile.....ProGlyAlaProAlaIleLysLysHisAlaMetLeuS 401
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
700 CA.....AGTATTTCACACTATGACACGAATGGTAGATACAGTGA 743
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
401 erAspLysArgHisValLeuThrArgAspSerAspGlyAsnVal..... 415
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
744 AAAGTCCGGCTCGGAGGAGTAACACGCCACGCG..... 782
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
416 .....AlaLeuTyrAspValLeuAlaAlaArgLysIleLysAspTyr 429
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
783 .....AGAGTATACCCCGCAGCAGCAGTGCAGCAGCTGT 813
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
429 rGlyLysArgIlePheGluGluValValAspGluAsnSerArgGlnValT 446
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
814 ACATACCTGGGTGG.....ATT 830
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
446 yrIleProSerTrpPheValValAspSerLysSerGlyMetLeuGlnIle 462
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
831 GAGATGATGAATA.....CAATGGAATCCACCAAGACCTGA 868
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
463 ThrLeuAspGluLeuAspAlaLeuSerSerTrpLeuSerLysAspAl 479
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
869 A...ATTATAAGGATGATGTTCACACAAAGCAAGTGTGTTTCACGAGAA 915
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
479 agLysPheAspAspAsnAspArgGluThrLysGlnIleIle.....V 493
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
916 CATCTCCGGGAGCGTCATTTGTGGCTTCTGATGCCCAATGCTCCCATTA 965
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
493 aLasnTyrGlyGlyMetLeuArgSer.....Leu 503
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
966 CTGATCTTCCCGACCTGGCCAAATCACCGACCTGAAGCGGGAATTCAT 1015
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
504 PheGluArgTrpProProCysLysMetThrAsnValAspAla..... 517
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
1016 CGGGGGCAGTCTCATTAATCTGACTTGGCAGCTCTGGGATGATTATG 1065
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
518 .....AlaAspAlaAspValG 524
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
1066 ACCATGGAACAGCTCACAAAGTATATCATTCGAATAAGTACAGTATTCTT 1115
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
524 InLysAlaThr.....LeuAsnPheIle 531
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
1116 GATCTCAGACAGCAAGTTCAATGAATCTTTCAGTGAATACTACTGCTCT 1165
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
532 SerLeuProGluHis.....ThrProLe 539
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
1166 CATCCCAAGGAAGCAACTCTGAGGAAGTCTTT.....TTGTTTAAAC 1209
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
539 uIleIleCysGluGlyAsnGlyArgProLeuTyrArgLeuLeuValGlyA 556
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
1210 CAGAAACACTACTTTTGA...AATGCACAGATCTTTTTCATTGCTATT 1256
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
556 spAlaGlyGluPheGluAlaAsnGluLeuAlaGlnIleAlaProMet 572
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
1257 CAGGCTGTGTAGTGAAGTCGATCTGAATTCAGAAATATCCACATTCGACG 1306
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
573 TrpValIleAspAlaIleGlu...ArgAsnGlnLeuProLysPheAsnLy 588
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
1307 AGTATCTTTGTTTATT.....CCPCCACAGACTCCGCCAGACACACCTTA 1350
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
588 smetProPheTyrLeuLeuProHisProSerThrAsnProLysGlnProL 605
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
1351 GTCCTGATGAACAGCTGCTGCT..... 1370
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
605 ysLysAspArgLeuSerAlaThrGluMetLeuGlnValLysLysValMet 621
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
```

```
1371 .....CC 1372
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
622 GluHisValTyrGluLysIleLeuSerThrAsnAspValGlySerIlePr 638
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
1373 TTGTCTTAATATTCATCAACAGCACCATT..... 1403
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
638 oLeuAsnGlnIleHisThrLysMetGluMetTyrCysAsnAspGlnArgL 655
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
1404 .....CTGGCATTCAC.....ATTATAAAATTTATGGAGGTGG 1439
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
655 euGluProAspMetAspLeuArgThrValLysHisLeuTyrTrpLysGln 671
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
1440 ATAGAGAACTCCAGCTG 1457
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
672 SerGlyGluLeuLeuLeu 677
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
```

seq_name: pir2:T21808

seq documentation block:

hypothetical protein F35G12.4a - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T21808

R:Chui, C.

submitted to the EMBL Data Library, October 1994

A:Reference number: Z19473

A:Accession: T21808

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-686 <WIL>

A:Cross-references: EMBL:Z46242; PIDN:CAA86335.1; GSPDB:GN00021; CESP:F35G12.4a

A:Experimental source: clone F35G12

C:Genetics:

A:Gene: CESP:F35G12.4a

A:Map position: 3

A:Introns: 62/3; 105/1; 135/3; 200/3; 269/3; 332/2; 390/1; 425/3; 458/3; 490/1; 542/2

alignment_scores:

Quality: 131.00 Length: 490

Ratio: 0.577 Gaps: 27

Percent Similarity: 46.327 Percent Identity: 21.429

alignment_block:

US-09-049-696-19 x T21808 ..

Align seg 1/1 to: T21808 from: 1 to: 686

```
60 GAGGAGCTGTCCAAATGACAGGAGTTTACAGACATATGCTTCAGATCA 109
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
211 GluLysIleMetLysLeuArgGlyHisThrAspAsnValArgAlaLeuVa 227
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
110 AGTTCAGAACAAATGGCTCATTTGATGCTTTTGGGGCCCTTTCATCAGGAA 159
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
227 lValAsnAspAspGlyThr.....ArgAlaLeuSerAlaGlys 240
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
160 ATGGAGCTGTC.....TTCAGCGTCCATCCAG 188
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
240 erAspAlaThrIleArgLeuTrpAspIleGlyGlnGlnArgCysIle... 255
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
193 CTTGAGAGTAAGGGATTAAACCTCCAGACACCCAGTGGATGAATGGCAC 238
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
256 .....AlaThrCysIleAlaHisGluGlyValTrr.....Th 267
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
239 AGTGATCGTGGACAGC.....ACCGTGGGAAAGGACA 270
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
267 rIeuGlnValAspSerSerPheThrThrValTyrSerAlaGlyLysAspL 284
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
271 CTTTGTGTTTTCATCAC.....TGGACACGACAGCCTCCCAATCCTT 314
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
284 ysMetValValLysThrProLeuTyrAspPheThrLysSerGlnLeuLeu 300
      ::::|::|::| ::::|::|::| ::::|::|::| ::::|::|::|
315 CTCCTGGGATCCAGTGGACAGAAGCAAGTGGCTTTGTAGTGGACAAAAA 364
```

```

301 pheLysGluGluAlaProValLysLys..... 309
365 CACCAAAATGGCTACCTCCAAATCCAGGCATGTGCTAAGGTGGCACTT 414
310 LeuLeuLeuSerGluLysAspAsnProValSerLeuTrpValGlyThrT 326
415 GGAAT.....TACAGTCTCCAGCAAGCTCACAACCTTGG 449
326 rPLysSerAspIleLysArgTrpSerIleArgProSerAlaGln..... 340
450 ACCCTGACTGTACCTCCCTGGCTGCTCAATGCTACCTGCTCCCTCCAATTAC 499
341 LeuSerIleGlyLysGluAspGlyProSer..... 351
500 AGTGACTTCCAAACGACAGACACCAAGCAATTTCCCGAGCCCTCTGG 549
352 ThrSerAsnAlaAsnHisSerValSerAlaSerSerProProV 367
550 TAGTTTATGCAATATTCCGCAAGGAGCCCTCCCAATTCTCAGGCGCAGT 599
367 alThrPheLysTyIleArg..... 373
600 GTCACAGCCCTGATGAATCAGTCAATGAAAGCAACAGTTTACCTTGAAC 649
374 ValLysAspGlnLysGlyGlnSerThrProGluLe 386
650 ACTGGATGAGGACAGCTGCTGATGCTACTAAGGATGACGGTGTCTACT 699
386 uValle.....ProGlyAlaProAlaIleLysLysHisAlaMetLys 401
700 CA.....AGCTATTTCACAACTTATCACACGAAATGATGATACAGTGTA 743
401 erAspLysArgHisValLeuThrArgAspSerAspGlyAsnVal..... 415
744 AAAGTGGCGGCTCTGGAGGATTAACGCGCCAGACGG..... 782
416 AlaLeuTyAspValLeuAlaAlaArgLysIleLysAspTy 429
783 .....AGAGTGATACCCAGCAGAGTGAGGACACTGT 813
429 rGlyLysArgIlePheGluGluValValAspGluAsnSerArgGlnValT 446
814 ACATACCTGGCTGG.....ATT 830
446 yIleProSerTrpPheValValAspSerLysSerGlyMetLeuGlnIle 462
831 GAGATGATGAATA.....CAATGGAATCCACCAAGACCTGA 868
463 ThrLeuAspGluLeuAspAlaLeuSerSerTrpLeuSerSerLysAspAl 479
869 A...ATTAAAGGATGATGTTTACACAGCAAGTGTGTTTCAGCAGAA 915
479 aGlyPheAspAspAsnAspArgGluThrLysGlnIle.....V 493
916 CATCTCTGGGAGGCTCATTTGTGGCTTCTGATGCCAAATGCTCCCAT 965
493 aIAsnTyrglyGlyMetMetLeuArgSer.....Leu 503
966 CTGATCTCTCCACCTGGCGCAATACCCAGCTGAGCGGCAATTC 1015
504 PheGluArgTrpProProCysLysMetThrAsnValAspAla..... 517
1016 CGGGGCGAGTCTCATTAACTGACTTGGACAGCTCTCTGGGATGATTATG 1065
518 .....AlaAspAlaAspValG 524
1066 ACCATGGAACAGCTCACAAGTATATCATTCGAATAGTACAAGTATTCTT 1115
524 InLysAlaThr.....LeuAsnPheIle 531
1116 GATCTCAGACAGAGTTCAATGAATCTCTTCAAGTGAATACCTACCTGCT 1165
532 SerLeuProGluHis.....ThrProLe 539
1166 CATCCCAAGCAAGCACTCTGAGGAAGTCTTT.....TTGTTTAAAC 1209
539 uIleleCysGluGlyAsnGlyArgProLeuTyArgLeuLeuValGlyA 556
1210 CAGAAAACATTACTTTTAA.....AATGGCACAGATCTTTTCATTGCTATT 1256
556 spAlaGlyLysGluPheGluAlaAsnGluLeuAlaGlnIleAlaProMet 572
1257 CAGCTGTGTATAAGCTCGATCTGAAATCAGAAATATCCACATTCGACG 1306
573 TrpValIleAspAlaIleGlu.....ArgAsnGlnLeuProLysPheAsnL 588
1307 AGTATCTTTGTTTATT.....CCTCCACAGACTCCCGCAGACACCTA 1350
588 smetProPheTyrlLeuLeuProHisProSerThrAsnProLysGlnProL 605
1351 GTCCGTGATGAAACGCTCTGCT 1370
605 ysLysAspArgLeuSerAla 611
seq_name: pir2:C81675
seq_documentation_block:
polymorphic membrane protein B/C family TC0694 [imported] - Chlamydia muridarum (stra
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: C81675
R:Read: T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Winn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: C81675
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1672 <TET>
A:Cross-references: GB:AE002338; GB:AE002160; NID:g7190724; PIDN:AAF39510.1; PID:g719
A:Experimental source: strain N199 (MoPn)
C:Genetics:
A:Gene: TC0694
alignment_scores:
Quality: 130.00 Length: 548
Ratio: 0.570 Gaps: 22
Percent Similarity: 41.606 Percent Identity: 19.161
alignment_block:
US-09-049-696-19 x C81675
Align seg 1/1 to: C81675 from: 1 to: 1672
9 GGTGCCATCATCCACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAACT 58
263 GlyAlaIleTyrlAlaThrGlyAspLeuThrIleSerAspSerGlnGluI 279
59 A.....GAGAGCTGTCCAAATGACAGGAGTTTA..... 89
279 eValPheSerValAsnLysAlaSerLysAspGlyGlyAlaIlePheAlaG 296
90 .....CAGACATATGCTTCAGATCAAGTTCAGAACT 122
296 luLysAsnValSerPheGluAsnIleThrThrLeuLysValGlnAsnAsn 312
123 GGCCTCATTCATGCTTTTGGGGCCCTTTTCATCAGAAATGAGAGCTGTCT 172
313 .....GlyAlaGluGluLysGlyGlyIleTyrl 323
173 TCACGCTCTCATCAGCTTGAGAGTAGGAGTTAACCTCCAGACAGCC 222
323 aSerGlyAspLeuSerIleGlnSerLysGlnSerLeuPheAsnSer. 339
```



```
223 AGTGGATGATGGGACAGTGTCTGTTGGACAGCAGCGTGGGAAGACACT 272
339 .....
273 TTGTTTCTTATCACTCGACACAGCAGCTCCCAAAATCCTTCCTGGGA 322
340 .....
323 TCCAGTGGACAGCAAGTGGCTTTGTAGTGACAAACACCAAAA 372
340 nThrSerLysGlnGlyGlyAlaLeuTyrIleGluGlyAsnValAsp 357
373 TGGCCTACCTCCAAATCCAGGCAATGCTAAGTTGGCACTTGGAAATAC 422
357 heLysAspLeuGluGluIleArgile.....LysTyr 367
423 AGTGTGAAGCAAGCTCAACACTTGACCTGACTGTACGTCCTCGGTGC 472
368 AsnLysSerGlyThrPheGluThrLysLysValThrLeuSerLeuProG 384
473 GTCCAATGCTACCTGCTCCAAATACAGTACTTCCAAAACGACAAAG 522
384 uAlaGlnThrAsnLysSerSerValThrAlaAlaSerGlnSerGlyPro 401
523 ACACGACCAATTCACAGCCCTCTGTA.....GTT 554
401 snThrThrProThrProThrProValThrAlaLysGlyGlyLeu 417
555 TATGCAATATTCGCAAGAGCCTCCCAATTCACAGGCGCAGTGTAC 604
418 TyrThrGluLysAsnLeuSerIleSer.....AsnIleTh 429
605 AGCCCTGATGAATCAGTGAATGGAACACAGTACCTTGAAGTACTGG 654
429 rGlyIleIleGluIleThrAsnAsnLysAlaThr.....A 441
655 ATATGGAGCAGTCT..... 671
441 spValGlyGlyAlaTyrValLysGlyThrLeuThrCysLysAspSer 457
672 .....GATGCTACTAAGGATGACGGTGT 694
458 HisArgLeuGlnPheGlnLysAsnSerSerGluLysLysGlyGlyLe 474
695 CTACTCAAGGTATTTCAACTTATGACACCAAGTGGTAGATACAGTAA 744
474 uTyr.....ThrGluAspThr.....IleT 481
745 AAGTGGGGCTCTGGAGGAGTTAACGCGCAGCAGCGAGAGTATACCC 794
481 hrLeuSerAsnLeuThrGlyLysThrLeuPheGlnGluAsnThrAlaLys 497
795 CAGCAGTGGAGCAGCTACATCTCGCTGGGATTCAGAAATGATGAAT 844
498 GluGluGlyGlyGlyLeuTyrIleGlnGly..... 507
845 ACAATGGAATCCACCAAGACCTGAAATTAATAGGATGAT..... 884
508 .....AspAspLysThrLeuT 513
885 .....GTTCAACACAGCAAGTGTGTTTC.....AGCAGACA 917
513 hrMetThrGlyLeuAspSerPheCysLeuIleAspAsnThrSerAlaThr 529
918 TCCTCGGAGGCTCATTTGTGGCTCTGATGTCCTCAATGCTCCCATACC 967
530 HisGlyGlyGlyAlaTyrValThrLysGluIleSerGlnThrTyrThrSe 546
968 TGATCTC.....TTCACACCTGGCCAAATCACCAGCTGAAGCGGAAA 1011
546 rAspValGluGluPhePro.....GlyIleThr.....ProV 557
```

```
1012 TTCACGGGGCAGTCTCATTAATCTGACTTGGACAGCTCCTGGGATGAT 1061
557 alHisGlyGluThrIleSerGlyAsnLysAlaThrGlyGlySerGly 573
1062 TATGACCATGAACAGCTCAACAAGTATATCATTCGAATAAGTACAGTAT 1111
574 GlyGlyValCysThrLysHisLeuValLeuSerAsnLeuGlnThrIleSe 590
1112 TCTTGATCTCAGAGACAAGTTCAATGAATCTCTTCAAGTGAATACTACTG 1161
590 rIleSer..... 592
1162 CTCTCATCCCAAGGAAGCAACTCTGAGGAAGTCTTTTGTGTTAAACCA 1211
592 ..... 592
1212 GAAACATTTACTTTTGAATAATGGCAGACAGATCTTTTCATTTGCTATCAGGC 1261
593 GluAsnPheAlaSerGluAsnGlyGly.....AlaCysThrCy 606
1262 TGTGTATAAGCTCGATCTGAAATCAGAAATATCCAACATTGCACAGTAT 1311
606 sProAspAsnPhe..... 610
1312 CTTTGTGTTATCTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAA 1361
611 .....ProAlaProThrAlaSerThrProSerThrAsnGln 622
1362 ACGTCTGCTCCT.....TGCTCTAATATTCATATCAAA 1393
623 ThrAlaAlaProLysAspLysAspPheLeuIleAspTyrValValSe 639
1394 CAGCACCATT.....CCTGGCATTCACA 1416
639 rThrThrIleAspLysAsnLysAlaThrLysLysGlyAlaGlyValTyrA 656
1417 TTTTAAATTTATGTGAAGTGGATGAGAGAACCTGCAGCTGCA 1460
656 laLysLysAlaLysLeuSerArgIleAspGluLeuAsnIleSer 670
seq_name: pir2:T30805
seq documentation_block:
dutt1 protein - mouse
N;Alternate names: transmembrane receptor protein Robol homolog
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T30805
R;Wu, M.C.; Lowe, N.; Fordham, R.; Rabbitts, P.
submitted to the EMBL Data Library, July 1998
A;Description: The mouse homologue of human DUTTL/H-robol gene: protein sequence and
A;Reference number: Z20879
A;Accession: T30805
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1612 <WUM>
A;Cross-references: EMBL:Y17793; NID:el329712; PID:el329713; PIDN:CAA76850.1
A;Experimental source: brain
C;Genetics:
A;Gene: dutt1
A;Map position: 16
```

```
alignment_scores:
Quality: 128.00 Length: 467
Ratio: 0.593 Gaps: 22
Percent Similarity: 46.253 Percent Identity: 20.771
alignment_block:
US-09-049-696-19 x T30805
Align seg 1/1 to: T30805 from: 1 to: 1612
```

228 ATGAATGGCACAGTGTGCTGGACAGCACCCTGGGAAAGGACACTTTGTT 277
::: ||| ||||| ||||| :::: |:::
423 ValAsnGlnThrValAlaValaspGlyThrLeu.....IleLe 435

278 TCTTATCACTGGNACAACGCAGCCCTCCCANAATCTCTCTGGGATCCCCA 327
::: ||| :::: |:::
435 userCysValalathrGlySerProAlaProThrIleLeutrp..... 449

328 GTGACAGAAGCAAGGTGGCTTTGTAGTGCGAAAAAACACCAAATGCC 377
::: ||| :::: |:::
450ArgLysAspGlyValLeuValserThrGlnAspSerArgIleLys 464

378 FACCTCCAATCCAGCGCATTT.....GCTAAGGTGGC... 410
||||: ||::: |:::
465 GlnLeuGlu.....SerGlyValLeuGlnIleArgTyraLaLysLeuGlyAs 480

411 ACTTGGAAATACAGTCTGCNACAAGAGCTCACAAAACCTTGACCTGACGTG 459
||| :::: |::: |:::
480 pThrGlyargtyrThrCysThrAlaserThrProSerGlyGluAlaThrt 497

460 TCACGTCC.....::: 467

497 rpSerAlaTyriLeuValGlnGluPheGlyValProValGlnProPro 513
::: |:::

468 CGTCGCTCCAATGCTACCTGCCTCCAATTACAGTGACTTCCAAAACGAA 517
||| :::: |::: |:::
514 ArgProThrAspProAsnLeuIleProSerAlaProSerLysProGluVa 530

518 CAAGGACACCAAGCAAAATCCCGACGCTCTGTGTAGTTTTATGCAAAATATTC 567
||| :::: |::: |:::
530 lThrAspValSerLysAsnThrValThrLeuSerTrpGlnProAsnLeuVa 547

568 GCNAGGAGCCTCCCA.....ATTCTCAGGCC..... 596
::::: ||||: |:::
547 snSerGlyAlaThrProThrSerTyriLeileGluAlaPheSerHisla 563

597AGTGCACACGACCTGATTGAATCAGTAGTAATCGAAAAACAGT 637
||| ::: ||||: |::: |:::
564 SerGlySerSerTrpGlnThrAlaAlaGluasnValLysThrGluThrPh 580

638 TACCTTGGAA.....CTACTGGATAATG 660
||: ||| :::: |:::
580 eAlaIleLysGlyLeuLysProAsnAlaIleTyriLeuPheLeuValLara 597

661 GACAGCGTCTGATCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCT 710
::: ||| :::: |:::
597 laalaasnalatyrGlyIleseraspProSerGlnIleSeraspProval 613

711 ACAACTTATGAC.....ACGAATGGTAGATACAGTGTAAAGT 748
||| ::: :::: |::: |:::
614 LysThrGlnAspValProProThrSerGlnGlyValAspHisLysGlnVa 630

749 G....CGGGCTCTGGGAGGNTT..... 767
||| ||| ||||: |::: |:::
630 lGlnArgGluLeuGlyAsnValValLeuHisLeuHisAsnProThrIleL 647

768AACGCGACGACGAGGAGTGTACCCACAGAGAT... 803
::: ||| :::: |:::
647 euSerSerSerValGluValHistTrpThrValAspGlnInSerGln 663

804GGAGCACTGTACATACCTGGCTGGATTGAGATGA 838
||| ||| |::: |:::
664 TyriLeGlnGlyTyriLysIleLeutyrrgProSerGlyAlaserHisGl 680

839 TGAATCAATATG.....AATCCA..... 857
||| :::: |::: |:::
680 yGluserGluTrpLeuValPheGluValArgThrProThrLysAsnServ 697

858CCAAGA 863
|||

697 alValIlePrnaspLeuArgLysGlyValAlasntyriGluIleLysAlaArg 713

864 CCTCAAAATTAATAGGATGATCTTCAACACGAAGTGTGTTTCAGCAG 913

```

714 PropPheAsnGluPheGlnGlyAlaAspSerGluIleLysPheAlaTy 730
    ||| :|||::: ::::: :|||::: |::|:::
914 AACATCTCGGGAGGTCAATTGTGGCTTCTGATGTCGCCAATGCTCCA 963
    :|| :|||::: :|||::: :|||::: :|||:::
730 sThr.....LeuGluGluAlaProSerAla... 738
    :|||::: :|||::: :|||::: :|||::: :|||:::
964 TACCTGATCTCTCCACACTGGCCAAATCACCCGACCTGAAGCGGGAAT 1013
    ||||| :|||::: :|||::: :|||::: :|||:::
739 .....ProProArgSerValThrValSerLysAsnAspGly 750
    :|||::: :|||::: :|||::: :|||::: :|||:::
1014 CACGGGGCAGCTCATTAATCTGCATGTGGACAGCTCTGGGATGATTA 1063
    :||| :|||::: :|||::: :|||::: :|||:::
751 AsnGlyThrAlaIleLeu..ValThrTrpGlnProProGluAspTh 766
    :|||::: :|||::: :|||::: :|||::: :|||:::
1064 TGACCATGGACAGCTCACAGTATATCATTCGAATAAGTACAAGTATTC 1113
    :|||::: :|||::: :|||::: :|||::: :|||:::
766 rGlnAsnGlyMetValGlnGluIyrLysValTrpCysLeuGlyAsn.... 781
    :|||::: :|||::: :|||::: :|||::: :|||:::
1114 TTGATCTCACAGACAAGTTCAATCAATCTCTTCAAGTGAATACT..... 1157
    :||| :|||::: :|||::: :|||::: :|||:::
782 ..GluThrLysTryHisIleAsnLysThrValAspGlySerThrPheSer 797
    :|||::: :|||::: :|||::: :|||::: :|||:::
1158 .....ACTGCTCTCATCCCCAAGAAGCCCAACTCTGAGGAAGTC... 1196
    :|||::: :|||::: :|||::: :|||::: :|||:::
798 ValValIleProSerLeuValProGlyIleargTyrSerValGluValAl 814
    :|||::: :|||::: :|||::: :|||::: :|||:::
1197 .....TTTTTGTTAAACCCAGAACACATTA 1221
    :|||::: :|||::: :|||::: :|||::: :|||:::
814 aAlaSerThrGlyAlaGlyProGlyValLysSerGluProGlnPheIleG 831
    :|||::: :|||::: :|||::: :|||::: :|||:::
1222 CWTTTGAANAATGGCACAGATCTTTTCATTCGTATTCAGGCTGTGTATAAG 1271
    :|||::: :|||::: :|||::: :|||::: :|||:::
831 InLeuAspSerHisGlyAsn.....ProValSerProGluAspGln 844
    :|||::: :|||::: :|||::: :|||::: :|||:::
1272 GTCGATCTGAATACGAATAATCCAAACATTCGACGAGTATCTTTGTATT 1321
    :|||::: :|||::: :|||::: :|||::: :|||:::
845 ValSerLeuAlaGlnGlnIleSerAspValValArgGlnProAlaPheIl 861
    :|||::: :|||::: :|||::: :|||::: :|||:::
1322 T 1322
    |
861 e 861

```

seq name: pir1:D54689

seq_documentation_block:
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type delta, splice form D precursor
N;Alternate names: MPTP delta type D
N;Contents: protein tyrosine phosphatase, receptor type delta, splice form A
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: D54689; A54689
R;Mizuno, K.; Hasegawa, K.; Katagiri, T.; Ogimoto, M.; Ichikawa, T.; Yakura, H.
Mol. Cell. Biol. 13, 5513-5523, 1993
A;Title: MPTP delta, a putative murine homolog of HPTP delta, is expressed in specialized
A;Reference number: A54689; MUID:93360986

A:Accession: D54689
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1691 <MI2>
A:Experimental source: Brain
A:Note: sequence inconsistent with nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:137486, NCBIPI:136537)
A:Accession: A54689
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-398,799-1691 <MI2>

a; Experimental source: brain
A; Note: sequence inconsistent with nucleotide translation
A; Note: sequence extracted from NCBI backbone (NCBIN:136522, NCBIPI:136524)
C; Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homolog
cgy
C; Keywords: alternative splicing; glycoprotein; phosphoprotein; phosphoric monoester

612 eAsnHisThrGlnMetAsnThrAlaLeu..... 621
835 ATGATGAATACAATCGAATCCACCA..... 860
622IleGlnrPHisProValAspThrPheGlyLeuGlnGly 635
861AGACCTGAATTAATAAGATGATGTCAA..... 890
636 TyrArgLeuLysPheGlyArgLysAspMetGluProLeuThrThrLeuG1 652
891CACACGCAAGTGTGTTTCACGACGACATCC.....TCGGGAG 927
653 upheSerGluLysGluAspHisPheThrAlaThrAspIleHisLysG1VA 669

891CACAGCAAGTGTGTTTCACGACATCC.....TCGGCAG 927
036 TTTATGdeU₁ysPheCYargLYsaspMetGUprOleUrnTrnLeuG1 655
036 upheserGluUysGLAspHisPheThrAlaTrnAspLHeHisValGUA 669
652

928 GCTCATTGTG GCT 941

[illegible]

669 laSerTvrValPheArgLeuSerAlaArgAsnLvsValGlvPheGlvGln 685

.. 54.1

942 TCATGATGTCCTCCAAATGCTCCCATACCTGATCTCTCTCCACCTGGCCAAAT 991
::: ||| ::: :::: |||::: ::||| |||
686 GluMetValLysGluIleSerValProGluGluIleProThrGlypHePr 702

992 CACCGACCTGANGGGGAAATTACGGGGCAGTCTCATTAACTGACTT 1041

702 oGlnAsnLeuHisSerGluGlyThrSerThrSerValGlnLeuSerT 719

1042 GGACAGCTCCTGGGGATGATTATGACCATGGAACAGCTCACAGTATATC 1091

719 rpGlnProValLeuAlaGluArgAsnGlyValIleThrLysTyr... 734

1092 ATTCGAATAAGTACAAGTATTCTTGATCTCAGAGACAAGTTCATGAATC 1141

1142 TCTTCAAGTGAATACTACTGCTCTCATCCCAAGGAAGCCAACTCTGAGG 1191

```

4
735 .....ThrLeuLeuTyArgAspIleAsn..... 742
      ::::: ::::: |||

```

1192 AAGTCTTTTGTTTAAACCAGAAAACATTACTTTTGAAAATGGCACAGAT 1241

743 ..ValProLeuProMetGluHisLeuIleValProAlaAspThrSer 758

1242 CTTTCATTGCTATTCAGGCTGTTGATAAGGTGCATCTCGAAATCAGAAAT 1291
:: :: :: ::
Thrgl v[eu]ysSerAsnTh 768

759 *metcinn* deu.....imgrylculysseraspin 768
1292 ATCCAACATTGCACGAGTATCTTTGTTTATTCCTCCACAGACTCGCCAG 1341

```

768 rThrTyraspValLysvalArgAlaHisThrSerLysArgArgGlyProt 785
      ... : ::::: |||   ::: |||

```

1342 AGACACCTAGT'.....CCTGATGAACGTCTGCTCCT 1373

785 yrSerProSerValGlnPheArgThrLeuProValAspGlnSerMetPhe 801

1374 TGTCCTAATATTTCATATCAACAGCACCATT 1403

```
seq name: p1r2:C54689
```

```
seq_documentation_block:
```

N: Alternate names: MPTP delta type B/C

N: contains: protein tyrosine phosphatase, receptor type delta, s
C: Species: Mus musculus (house mouse)
C: Date: 25-Apr-1995 #sequence_revision 19-May-1995 #text_change

C.; Accession: C54689; B54689
R.; Mizunō, K.; Hasegawa, K.; Katagiri, T.; Ogimoto, M.; Ichikawa,

Mol. Cell. Biol. 13, 5513-5523, 1993
A>Title: MPTP delta, a putative murine homolog of HPTP delta, is
A-reference number: 154580. MIM: 60260096

A;Reference number: A54689; MUID:93360986
A;Accession: C54689

624 ANTGGAAACAGTTACCTTGGAACTACTGATATATGGACAGGTGCTGA 673
|||||::: ::: ||| |||
750 AsnGlyGluProLysSerAlaMetLeuLysAspValMetLeu...AlaAs 765
674 TGCTACTAAGGATGACGGTGCTACTCAA 702
|||| ::: |||||
765 pAlaGlnAspMetIleIleSerGlyLeuGlnProGluThrSerTyrSerL 782
703 GGTATTTCACAACTTATGACACGAATGCTAGATACAGTGTA 743
|||||::: |||||::: |||
782 euThrValThrAlaTyrThrThrLysGlyAspAlaArgSerLysPro 798
744 AAA...GTCCGGCTCTGGAGAGATTACCCAGCAGCCAGGAGATCAT 790
||| ||| ::: |||||::: ||| |||||
799 LysLeuValSerThrThrGlySerValProGlyLysProArgLeuValI 815
791 A.....CCCCAGCAGTGGAGCACTGTACATACCTGGCTGGATTGAGA 834
| ||||::: |||||
815 eAsnHisThrGlnMetAsnThrAlaLeu 824
835 ATGATGAAATACAAATGGAATCCACCA 860
|||||::: |||||
825IleGlnTrpHisProProValAspThrPheGlyLeuGlnGly 838
861 ..AGACCTCAAATTAATAAGGATGATGTTCAA 890
||| ||| |||||::: |||||
839 TyrArgLeuLysPheGlyArgLysAspMetGluProLeuThrThrLeucl 855
891CACAAACAAGTGTGTTTCAGCAGAACATCCTCGGGAG 927
|||::: |||||::: |||||::: |||||
855 uPheSerGluLysGluAspHisPheThrAlaThrAspIleHisLysGlyA 872
928 GCTCATTTGTGGCT 941
::: |||||::: |||||
872 lAserTyrValPheArgLeuSerAlaArgAsnLysValGlyPheGlyGlu 888
942 TCTGATCTCCAAATGCTCCCATCTACCTGATCTCTCCCACTGGCCAAAT 991
::: |||::: |||||::: ||||| |||
889 GluMetValLysGluIleSerValProGluGluIleProThrGlyPhePr 905
992 CACCGACTCAAGCGCAAAATTCACGGGGCAGCTCTCATTATCTCAGTT 1041
::: |||::: |||||::: |||||::: |||||
905 oGlnAsnLeuHisSerGluGlyThrThrSerThrSerValGlnMetSerT 922
1042 GGACAGCTCTCTGGGAGTATTATGACCATGGAAACGCTCACAAGTATATC 1091
|| |||::: |||||::: |||||
922 rpGlnProProValLeuAlaGluArgAsnGlyValIleThrLysTyr... 937
1092 ATTCGAATAAGTACAAATATTCTTGATCTCAGACAGAGTTCAATGAATC 1141
937 937
1142 TCCTCAAGTGAATACTACTGCTCTCATCCCAAGAGCAACTCTGAGG 1191
::: |||||::: ||||| |||
938ThrLeuLeuTyrArgAspIleAsn..... 945
1192 AAGTCTTTTGTATTAACCAAGAAACAACTATTCTTTGAAAATGGCACAGAT 1241
|||::: |||||::: |||||
946 ..ValProLeuLeuProMetGluHisLeuIleValProAlaAspThrSer 961
1242 CTTTTCATTGCTATTACGGCTGTTGATAAGGTCGATCTGAAATCAGAAAT 1291
::: |||::: |||||::: |||||
962 MetThrLeu.....ThrGlyLeuLysSerAspTh 971
1292 ATCCAAACATTGCAGGAGTATCTTTGTTTATTCTCCACAGACTCCGCCAG 1341
|||::: |||||
971 rThrTyrAspValLysValArgAlaHisThrSerLysArgArgGlyProT 988
1342 AGACACCTAGTCCTGATGAACGCTCTGCTCCT 1373
::: |||||
988 ySerProSerValGlnPheArgThrLeuProValAspGlnSerMetPhe 1004
1374 TCTCTCAATATTTCATATCAACAGCACCATT 1403

1746CysThrCysArgAsnLeuHisGlnSerAsnThrSerArgAlaGluL 1761
474TCAAGTCTACCTGCTCCAAATACAGTCACTTCCAAAACG 515
1761 euLeuValThrGluAlaProSerLysProIleThrValThrValGluGlu 1777
516 AACAGGACACCAAGAAATCCCGACGCTCTGGTGTATGTCAGAAATAT 565
1778 GlnArgSerGlnSerValArgProGlyAlaAspValThrPheIleCysTh 1794
566 TCGCCAGAGAGCTCCCAATCTCAGGGCCAGTGTACACCCCTGATTG 615
1794 rAlaLysSerLysSerProAlaTyThrLeuValTrpThrArgLeu.... 1809
616 NATCAGTGAATGGAACACAGTCTACCTTGGAACTACTGGATAATGGAGCA 665
1810HisAsnGlyLysLeuProThrArgAlaMetAspPheAsnGlyIle 1824
666 GGTCTGTATGCTACT.....AAGGATACGGTGTCTAC..... 698
1825 LeuThrIleArgAsnValGlnLeuSerAspAlaGlyThrTyValCysTh 1841
699TCAAGGTATTTCACAACTTATGACACGAATGGTAGATACAGTAA 744
1841 rGlySerAsnMetPheAlaMet.....AspGlnGlyThrAlaThrLeuH 1856
745 AAGTCCGGGCTCTGGAGGAGTAAACGACGACGAGGAGTATACCC 794
1856 isValGlnAla..... 1859
795 CAGCAGTGGAGCACTGTACATACCTGGCTGGATTGACATGATGAAT 844
1860SerGlyThrLeuSerAlaPro.....Valva 1868
845 ACAATGGAATCACCACAGACCTGAAATTAATGAAGATGATGTTCAACACA 894
1868 lserIleHisProProGlnLeu.....ThrValGlnProG 1880
895 AGCAAGTGTGTTACAGACAGACATCCTCGGAGGCTCATTTGTGGCTTCT 944
1880 lYglnLeuAlaGluPheArgCysSerAlaThrGlySer..... 1892
945 GATGTCCCAAAATGCTCCCACTCATCTC.....TTCCACACC 982
1893ProThrProThrLeuGluTrpThrGlyGlyProG1 1904
983 TGGCCAAATCACCACCTGAAGCGGAAATTCACGGGGCAGTCTCATTA 1032
1904 yGlyGlnLeu...ProAlaLysAlaGlnIleHisGlyGly...IleLeuA 1919
1033 ATCTGACTGGACAGCTCTCGGGGATGATTATGAC..... 1067
1919 rgLeuProAlaValGluProThrAspGlnAlaGlnIleThrLeuCysArgAla 1935
1068 CATGGAACAGCTCACAAGTATATCATTCGAATTAAGTCAAGTATCTTGA 1117
1936 HisSerSerAlaGlyGlnGlnValAlaArg.....AlaValLeuHi 1949
1118 TCTCAGACACAGTTCATGAATCTCTTCAAGTGAAT..... 1154
1949 sValHisGlyGlyGlyProArgValGlnValSerProGluArgThrG 1966
1155ACTACTGTCTCT..... 1166
1966 lnValHisAlaGlyArgThrValArgLeuTyrcysArgAlaAlaGlyVal 1982
1167ATCCCAAAGA 1177
1983 ProSerAlaThrIleThrTrpArgLysGluGlyGlySerLeuProProG1 1999
1178 AGCCAACTCTGAGAA.....GTCTTTTGTGTTAAACCAAGAAACA 1218
1999 nAlaArgSerGluArgThrAspIleAlaThrLeuLeuIleProAlaIle 2016

1219 TTACTTTTGAATGCGACAGATCTTTTCAATGCTATTCAGGCTGTTGAT 1268
2016 hrThrAlaAspAlaGlyPheTyLeuCysValAlaThrSerProAlaGly 2032
1269 AAGTCCATCTGAAATCAGAAATATCCACATTCACGAGTATCTTTGTT 1318
2033 ThrAlaGlnAlaArgMetGlnValVal...ValLeuSerAlaSerAspAl 2048
1319 TATTCTCCACAGACTCCGCCAGACACCTAGTCTGTGATGAACGCTCTG 1368
2048 aserProGlyValLysIleGluSerSerProSerValThrGluG 2065
1369 CTCCTTCTCTAATATTCATATCAACAGCACCATTCTCTGCATTCACATT 1418
2065 ly.....GlnThrLeuAspLeuAsnCysValValAlaGlySerAlaHis 2079
1419 TTAATAATTATGTGG 1433
2080 AlaGlnValThrTrp 2084
seq_name: pir2:A56178
seq_documentation_block:
protein:tyrosine-phosphatase (EC 3.1.3.48), receptor type delta precursor - human
N:Alternate names: protein-tyrosine-phosphatase BPTP-2
C:Species: Homo sapiens (man)
C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 21-Jan-2000
C:Accession: A56178; S12052; B44929
J: Pulido, R.; Krueger, N.X.; Serra-Pages, C.; Saito, H.; Streuli, M.
J. Biol. Chem. 270, 6722-6728, 1995
A:title: Molecular characterization of the human transmembrane protein-tyrosine phosphatase delta isoforms.
A:Reference number: A56178; MUID:95204468
A:Accession: A56178
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1912 <PUL>
A:Cross-references: GB:J38929; NID:9755652; PIDN:AAC41749.1; PID:9755653
R: Krueger, N.X.; Streuli, M.; Saito, H.
EMBO J. 9, 3241-3252, 1990
A:title: Structural diversity and evolution of human receptor-like protein tyrosine phosphatase
A:Reference number: S12049; MUID:91006018
A:Accession: S12052
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 390-1912 <KRU>
A:Cross-references: GB:X54133; NID:935789; PIDN:CAA38068.1; PID:935790
A:Note: the sequence from Fig. 5B is inconsistent with that from Fig. 5A in having 56
R: Adachi, M.; Sekiya, M.; Arimura, Y.; Takekawa, M.; Itoh, F.; Hinoda, Y.; Imai, K.;
Cancer Res. 52, 737-740, 1992
A:title: Protein-tyrosine phosphatase expression in pre-B cell NALM-6.
A:Reference number: A44929; MUID:92119637
A:Accession: B44929
A:Molecule type: mRNA
A:Residues: 1756-1804, 'C', 1806-1845 <ADA>
A:Cross-references: GB:S78086; NID:9243545; PIDN:AAB21147.1; PID:9243546
A:Experimental source: pre-B cell NALM-6
A:Note: sequence extracted from NCBI backbone (NCBI:78086, NCBI:78087)
A:Note: the authors did not report the entire codon for residue 90
C:Genetics:
A:Gene: GDB:PTPRD
A:Cross-references: GDB:131384; OMIM:601598
A:Map position: 9p24-9p24
C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homolog
C:Keywords: glycoprotein; phosphoprotein; phosphoric monoester hydrolase; transmembr
F:38-100/Domain: immunoglobulin homology <IMM1>
F:140-209/Domain: immunoglobulin homology <IMM2>
F:250-304/Domain: immunoglobulin homology <IMM3>
F:711-811/Domain: fibronectin type III repeat homology <3FR>
F:1293-1912/Domain: leukocyte common antigen cytosolic domain homology <LAC>
F:1669-1892/Domain: protein-tyrosine-phosphatase homology <PTP2>

808 CACTGTACATACCTGGCTGGATTGAGAATGATGAATACAATGGAATCCA 857

A; Cross-references: EMBL: X640

A; Cross-references: EMBL: X640

seq_documentation_block:
titin, cardiac muscle [validated] - human
N:Alternate names: Connectin
N:Contents: serine/threonine-specific protein kinase (EC 2.7.1.-)
C:Species: Homo sapiens (man)
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Sep-2000
C:Accession: I38344; I38345; S20898; S20897; S20899; S63665; S37393
R:Label: S.; Kolmerer, B.
Science 270, 293-296, 1995
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330
A:Accession: I38344
A:Status: nucleic acid sequence not shown; translation not shown; translated from GS/1
A:Molecule type: mRNA
A:Residues: 1-26926 <LAB1>
A:Cross-references: EMBL:X90568; NID:g1017424; PID:g1017425
R:Musco, G.; Tziatzios, C.; Schnuck, P.; Pastore, A.
Biochemistry 34, 553-561, 1995
A:Title: Dissecting titin into its structural motifs: identification of an alpha-helical
A:Reference number: I38345; MUID:95119041
A:Accession: I38345
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1977-2014 <MUSX>
A:Cross-references: EMBL:X93270; NID:g602579; PIDN:CAA58243.1; PID:g602580
A:Note: conformation and properties are reported for a synthetic peptide corresponding
E:Label: S.; Gauthel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A:Title: Towards a molecular understanding of titin.
A:Reference number: S20897; MUID:92258380
A:Accession: S20898
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 13597-14200, 'I', 14202-14696 <LAB2>
A:Cross-references: EMBL:X64698; NID:g37192; PIDN:CAA45939.1; PID:g37193
A:Accession: S20897
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 16330-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB3>
A:Cross-references: EMBL:X64699; NID:g37190; PIDN:CAA45940.1; PID:g37191

[illegible]

seq name: pir2:B35621

seq_documentation_block:
spore germination protein 270-11 - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 29-Oct-1999
C:Accession: B35621
R:Giorda, R.; Ohmachi, T.; Shaw, D.R.; Ennis, H.L.
Biochemistry 29, 7264-7269, 1990
A:Title: A shared internal threonine-glutamic acid-threonine-proline repeat defines a family of proteins
A:Reference number: A35621; MUID:91002566
A:Accession: B35621
A:Molecule type: DNA
A:Residues: 1-532 <GI2>
A:Cross-references: GB:J029116; GB:M33862; NID:G924278; PIDN:AAA73632.1; PTD:gl67881

alignment_scores:		
Quality:	119.50	Length: 433
Ratio:	0.604	Gaps: 18
percent similarity:	45.727	Percent Identity: 20.785

```

alignment_block:
  US-09-049-696-19 x B35621      ..
  Align seg 1/1 to: B35621 from: 1 to: 532

243 ATCGTGGACAGCACCGTGGGAAAGGACACATTGTTCTTATCACCTGGAC 292
   ::::::::::||| ::::::::::| ::::::::::| ::::::::::| ::::::::::|
13 LeuIleSerAlaThrPheAlaAsnAsnAla...PheIleValHisTrpAs 28
   ::::::::::| ::::::::::| ::::::::::| ::::::::::| ::::::::::|

293 AAGCAGACCTCCCAAAATCCTCTCTGGGATCCAGGTGGACGAGAGCAAG 342
   ::::::::::| ::::::::::| ::::::::::| ::::::::::| ::::::::::|
28 nSerAspSerIleSerLysLeuThrGlyGlnIleGlyAspThrIleS 45
   ::::::::::| ::::::::::| ::::::::::| ::::::::::| ::::::::::|

343 GTGGCTTTGTAGTGGACAAACAAACACAAA..... 371
   ::::::::::| ::::::::::| ::::::::::| ::::::::::| ::::::::::|
45 erPheTyThrSerAspGlyAsnSerHisAspValLysSerSerAspGly 61
   ::::::::::| ::::::::::| ::::::::::| ::::::::::| ::::::::::|

372 .....ATGGCCTACTCTCCAAATCCCAAGGATGCC 400
   ::::::::::| ::::::::::| ::::::::::| ::::::::::| ::::::::::|

62 SerValSerSerValPheSerGlySerLeuThrAsnProGlyIlePh 78
   ::::::::::| ::::::::::| ::::::::::| ::::::::::| ::::::::::|

401 TAAGGTT.....GCACCTGGAAATACAGTCTGCAAGCAA 435
   ||||| ::::::::::| ::::::::::| ::::::::::| ::::::::::|
78 eLysValThrLeuThrLysGluGlyAsnIleGluPheThr.....S 92
   ::::::::::| ::::::::::| ::::::::::| ::::::::::| ::::::::::|

436 GCTCAACAACTTGACCCCTGACTCTCAGGTCGCGTGCCTCAATGCTACC 485
   ||||| ::::::::::| ::::::::::| ::::::::::| ::::::::::| ::::::::::|
92 erSerTyAspGluGlyLeuSerAlaThrIleValValSerSerGlyGly 108
   ::::::::::| ::::::::::| ::::::::::| ::::::::::| ::::::::::|

486 CTGCCTCCAATTCACAGTACTCCAAAACAAAGGACACACAAATTT 535
   ||||| ::::::::::| ::::::::::| ::::::::::| ::::::::::| ::::::::::|
109 GlnIleProlleThrThrThrSerSerThrThrThrAspGlySerSerTh 125
   ::::::::::| ::::::::::| ::::::::::| ::::::::::| ::::::::::|

536 CCCACGACCTCTGTGAGTTTATGCAAAATATTCGCAAGAGGCTCCCCAA 585
   ||||| ::::::::::| ::::::::::| ::::::::::| ::::::::::| ::::::::::|
125 rProSer.....ThrProt 130
   ::::::::::| ::::::::::| ::::::::::| ::::::::::| ::::::::::|

```

seq_name: pir2:T02759

seq_documentation_block:
hypothetical protein - Acinetobacter sp. (strain ADP1) (fragment)
C:Species: Acinetobacter sp.
A:Variety: strain ADP1

C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
C:Accession: T02759

R:Segura, A.; Ornstun, N.L.
submitted to the EMBL Data Library, June 1997

A:Description: P2R153.

A:Reference number: 214724

A:Accession: T02759

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-918 <SEG>

A:Cross-references: EMBL:AF011339; NID:g2286203; PIDN:AAC27114.1; PID:g2286204

A:Experimental source: strain ADP1

alignment_scores:
Quality: 119.50 Length: 537
Ratio: 0.556 Gaps: 21
Percent Similarity: 40.037 Percent Identity: 20.484

alignment_block:

US-09-049-696-19 x T02759 ..

Align seg 1/1 to: T02759 from: 1 to: 918

```
18 ATCCACAGTCGCTTGGGCCCTCGCAGCTCAAGAACTAGAGGAGCT 67
   ||| |||:||||| |||:|||||
47 IleThrThrAlaThrAspProAla..... 55
68 GTCCAAATGACAGGAGTTACAGATATGCTTCAGATCAAGTTTCAGA 117
   ||| |||:||||| |||:|||||
56 .....GlyAsnThrSerAlaAlaSerAspAlaLeu.... 65
118 ACAATGGCCTCATGTATGCTTTTGGGCCCTTTCATCAGAAATGAGCT 167
   ||| |||:||||| |||:|||||
66 ..AsnPheThrValAspSerSerAsnValValSerIleAsnLysAla 81
168 GTCTCTCAGCGTCCATCCAGCTTGAGAGTAAGGATTACCCCTCCAGAA 217
   |||:||||| |||:|||||
82 IleAspAspAlaGly.....SerLysThrGlyAspLeuAlaAs 94
218 CAGCCAGTGGATGAATGGCACAGTGTGTCGACACCGCTGGGAAAGG 267
   |||:||||| |||:|||||
94 nasnAlaAlaThrAspAspThr.....ThrProThrLeuValGlyThrG 109
268 ACACCTTTGTTTCTTATCACCCTGGACACGCGCTCCCAAAATCCTTCTC 317
   |||:||||| |||:|||||
109 lyThrValGlyAlaVal.....ValSerIle 117
318 TGGGATCCAGTGGCAGAGAAGGCTGCTTTGTAGTGGCAAAAACAC 367
   |||:||||| |||:|||||
118 SerValAspGlyGlyAlaValValGlySerAlaValValAspSerAsn.. 133
368 CAAATGGCTACCTCCAAATCCAGGCATTTGCTAAGGTTGGCAGCTTGA 417
   |||:||||| |||:|||||
134 .....GlyAsnTrp 137
418 AATACACTGTGCAGCAAGCTCACAACCTTGACCTGACTGTGACGTCC 467
   |||:||||| |||:|||||
137 erThrThrLeuProSerGlnSerGluGly...ThrHisSerThrThrGly 152
468 CGTGGCGTCAATGCT..... 482
153 ThrAlaSerAsnAlaAlaGlyThrGlnGlyThrAlaSerPheThrLeuTh 169
483 .....ACCTGCTCCCAATTACAGTGAAGTTCACCAACGAAACAA 522
   ||| |||:||||| |||:|||||
169 rIleAspThrThrAlaProAspValProSerIleGlyGlnValAsnAspA 186
```

Tue Apr 2 09:39:54 2002

```

1144 TTCAAGTGAATACTACTGCTCTCATCCCAAGCAAGCCAACTCTGAGGAA 1193
      ::::: :::::
468 hrLysValThrAlaGluGlyLeu..... 475
1194 GTCTTTTGTTRAACCAAGAAACATTACTTTGAAAAATGGCACAGATCT 1243
      ||| ||| |||
476 .....ThrProGlyAsnTyrArgPheAspPheThrGlyGlyse 488
1244 TTTCATTGCTATTGAGGCTGTGATAGGTGATGATGAAATCAGAAATAT 1293
      ::::: ::::: ||::: ||:::
488 rLeuIleGlyLeuGlyThrSerIleLysAlaAspLeuGlnLeuThrThrg 505
1294 CCAACATTGCA 1304
      ::||| |||
505 lnAsnThrAla 508

```



```
161 TGGAGCTGCTCTCAGCGCTCCATCAGCTCAGAGTAAGGATTAAACC 210
: |||:|||||:
671 rLysAlaThrGlnAsnGlyLysLeuAsnValThrPheAsnGlyLysSerA 688
: |||:|||||:
211 TCAGAACACAGCTGGATGAATGGCACAGTGATCGTG.....GACAGC 254
: |||:|||||:
688 spGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAsnGlyAspLeu 704
: |||:|||||:
255 ACCGTGGGAAAGACACTTTGTTCTTATCCTCGGACAAAGCCAGCCVCC 304
: |||:|||||:
705 AsnValGluLysGlyThrLeuPheLeuSerGlyArgProThrProHisAl 721
: |||:|||||:
305 CCAATCCTCTCTCGGATCCCGAGTGAGCAAGCAAGTGGCTTTGTA. 353
: |||:|||||:
721 aArgAspIleAlaGlyIleSerSerThrLysLysAspProHisPheThrG 738
: |||:|||||:
354 .....GTGGACAAAACACCAA 371
: |||:|||||:
738 luAsnAsnGluValValGluAspAspTrpIleAsnArgAsnPheLys 754
: |||:|||||:
372 ATGGCTACTCTCAATCCAGCATTTGCTAAGTTGGCAGCTTGGAAATA 421
: |||:|||||:
755 AlaThrThrMetAsnValThrGlyAsnAla..... 764
: |||:|||||:
422 CAGTCTGCAAGCAAGCTCAAACTTGCACCTGACTGTACGTCCTCCGTG 471
: |||:|||||:
765 SerLeuTyrSerGlyArgAsnValAlaAsnIleThrSerAsnIleThrA 781
: |||:|||||:
472 GTGCCAAT.....GCTACCTGCTCCCAATT 497
: |||:|||||:
781 laSerAsnAsnAlaGlnValHisIleGlyTyrLysThrGlyAspThrVal 797
: |||:|||||:
498 ACAGTGACTTCCAAA..... 512
: |||:|||||:
798 CysValArgSerAspTyrThrGlyTyrValThrCysHisAsnSerAsnLe 814
: |||:|||||:
513 .ACGACAAGGACAGCAGCAAAATTC...CCAGCCCTCTGTTAGTTATG 558
: |||:|||||:
814 userGluLysAlaLeuAsnSerPheAsnProThrAsnLeuArgGlyAsnV 831
: |||:|||||:
559 CAATATTCCCAAGGAGCTCCCAATTC...AGGGCAGTGTCACA 605
: |||:|||||:
831 alAsnLeuThrGluAsnAlaSerPheThrLeuGlyLysAlaAsnLeuPhe 847
: |||:|||||:
606 GCCTGATTGAATCAGTGAATGAAACAGTGTACCTTGAA..... 647
: |||:|||||:
848 GlyThrIleGlnSerIleGlyThrSerGlnValAsnLeuLysGluAsnSe 864
: |||:|||||:
648 .....CTACTGATA 657
: |||:|||||:
864 rHisTrpHisLeuThrGlyAsnSerAsnValAsnGlnLeuAsnLeuThrA 881
: |||:|||||:
658 ATGGA.....GCAGGTGCTGATGCTACTAAGGATGACGGT 692
: |||:|||||:
881 snGlyHisIleHisLeuAsnAlaGlnAsnAspAlaAsnLys..... 894
: |||:|||||:
693 GTCTACTCAAGGATTTTCACACTTATGACAGGAATGGTAGTACAGTGT 742
: |||:|||||:
895 .....ValThrThrTyrAsnThr.....Le 901
: |||:|||||:
743 AAAAGTCGGGCTCTGGAGAGCTTAACGACGACGAGGAGTGTATC 792
: |||:|||||:
901 uThrValAsnSerLeuSerGly..... 908
: |||:|||||:
793 CCCAGCAGATGGAGCACTGTACATACCTGGCTGGATTGAGATGAA 842
: |||:|||||:
909 .....AsnGlySerPheTyrTyr.....TrpValAspPheThrAsn 920
: |||:|||||:
843 ATCAATGGATCCACCAAGCTGAATTAATGAAGTATGTTCAACA 892
: |||:|||||:
921 AsnLysSerAsn.....LysValValValAsnLys..... 930
: |||:|||||:
893 CAAGCAAGTGTGTTTCAGCAGAACATCCTCGGAGGCTCAITTTGTGCTT 942
```

```
931 .....|||:|||||:
: |||:|||||:SerAlaThrGlyAsnPheThrLeu. 938
943 CTGATGTCCCAAAATGCTCCCATACCTGATCTCTCCACACCTGGCCAAATC 992
: |||:|||||:
939 .....GlnVal 940
: |||:|||||:
993 ACCGACCTGAAGCGGAAATTCACGGGGGCGAGTCTCATT..... 1031
: |||:|||||:
941 AlaAspLysThrGlyGluProAsnHisAsnGluLeuThrLeuPheAspAl 957
: |||:|||||:
1032 .....AATCTGACTTGGACAGCTCCTCTGGGATGATT 1062
: |||:|||||:
957 aserAsnAlaThrArgAsnAsnLeuGluValThrLeuAlaAsnGlySerV 974
: |||:|||||:
1063 ATGACCATGGAAACAGCTCACAAATATATCATTCGAATAAGTACAAGTATT 1112
: |||:|||||:
974 alaAspArgGly...AlaTrpLysTyrLysLeuArgAsnValAsnGlyArg 989
: |||:|||||:
1113 CTTGATCTCAGA.....GACAAGTTCAATGAATCTTCTCAAGT 1150
: |||:|||||:
990 TyrAspLeuTyrAsnProGluValGluLysArgAsnGlnThr.....Va 1004
: |||:|||||:
1151 CAATACTACTCTCTCATCCCAAGGAAGCAACTCTGAGGAAGTCTTTT 1200
: |||:|||||:
1004 lAspThrThr..... 1007
: |||:|||||:
1201 TGTTTAAACCAAGAAACATTACTTTGAAATGGCACAGATCTTTTCATT 1250
: |||:|||||:
1008 .....AsnIleThrThrProAsnAsp..... 1014
: |||:|||||:
1251 GCTATTACAGCTGTGTGATAGGTGCTGAATCAGAAATCAGAAATATCAACAT 1300
: |||:|||||:
1015 ...IleGlnAla...AspAlaProSerAlaGlnSerAsnGluGluI 1029
: |||:|||||:
1301 TGCAGCAGTATCTTTGTTTATTCTCTCCACAGACTCCG..... 1337
: |||:|||||:
1029 eAlaArgValGluThrProValProProAlaProAlaThrGluSerA 1046
: |||:|||||:
1338 .....CCAGACACCTAGTCTCTGATGAACGTCGTCT 1370
: |||:|||||:
1046 laIleAlaSerGluGlnProGluThr...ArgProAlaGluThrAlaGln 1061
: |||:|||||:
1371 CCT 1373
: |||:|||||:
1062 Pro 1062
```

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US95-10661A-6

seq_documentation_block:

; Sequence 6, Application PC/TUS9510661A

; GENERAL INFORMATION:

; APPLICANT: Washington University, et al.

; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert

; STREET: 4 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: United States

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/10661A

; FILING DATE: 16-AUG-1995

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/296,791

```

; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: FP-59941/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; PCT-US95-10661A-6

alignment_scores:
  Quality: 117.50      Length: 551
  Ratio: 0.500        Gaps: 28
  Percent Similarity: 42.650      Percent Identity: 22.323

alignment_block:
US-09-049-696-19 x PCT-US95-10661A-6
Align seg 1/1 to: PCT-US95-10661A-6 from: 1 to: 1848

3 CAAAGTGGTGCATCATCCACAGCTGCCTTTGGGGCCCTCTGCAGCTCA 52
||||| : : : : : ||| : : : : : ||| : : : : :
607 GlnAspAsnArgSerTyrThrLeuLysLysGlyAlaSerThrArgSe 623
||||| : : : : : ||| : : : : : ||| : : : : :
53 AGAAGTACAGAGCTGTCACAAATGACA.....G 81
||||| : : : : : ||| : : : : : ||| : : : : :
623 rGluLeuProGlnAsnSerGlyGluSerAsnGluAsnTrpLeuTyrMetG 640
|| : : : : : ||| : : : : : ||| : : : : :
640 lYArgThrSerAspAlaAlaLysArgAsnValMetAsnHisLeuAsn 656
120 .....AATGGCCTCATGATGCTTTTGGGGCCCTTCATCAGGAAA 160
||||| : : : : : ||| : : : : : ||| : : : : :
657 GluArgMetAsnGlyPheAsnGlyTyrPheGly.....GluGluGluTh 671
161 TGGAGCTGCTCTCAGCCCTCCATCCAGCTTGAGAGTAAGGATTAACCC 210
||||| : : : : : ||| : : : : : ||| : : : : :
671 rLysAlaThrGlnAsnGlyLysLeuAsnValThrPheAsnGlyLysSerA 688
||||| : : : : : ||| : : : : : ||| : : : : :
211 TCCAGAACAGCAGTGGATGAATGCACAGTGCATCGTG.....GACAGC 254
||||| : : : : : ||| : : : : : ||| : : : : :
688 spGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAsnGlyAspLeu 704
255 ACCGTGGGAAGGACACATTTGCTTATACCTGGACACAGCCGCCCTCC 304
||||| : : : : : ||| : : : : : ||| : : : : :
705 AsnValGluLysGlyThrLeuPheLeuSerGlyArgProThrProHisAl 721
305 CCAATCCTCTCTGGGATCCAGCTGGACAGAACGAGTGCCTTTGTA. 353
||||| : : : : : ||| : : : : : ||| : : : : :
721 aArgAspIleAlaGlyIleSerSerThrLysLysAspProHisPheThrG 738
354 .....GTGGACAAAACACACAAA 371
738 luAsnAsnGluValValGluAspAspTrpIleAsnArgAsnPhelLys 754
372 ATGGCCTACCTCCAAATCCAGGATTCCTAAGGTTGGCACTTGGAAATA 421
||||| : : : : : ||| : : : : : ||| : : : : :
755 AlaThrThrMetAsnValThrGlyAsnAla..... 764
422 CAGTGTGCAAGACGCTCACAACTTGACCTGACTGCTACGTCGCCGTG 471
||||| : : : : : ||| : : : : : ||| : : : : :
765 .SerLeuTyrSerGlyArgAsnValAlaAsnIleThrSerAsnIleThrA 781
472 CGTCCAAT.....GCTACCTGCTGCCCAATT 497

```

```

||||| ||| : : :
781 laSerAsnAsnAlaGlnValHisIleGlyTyrLysThrGlyAspThrVal 797
498 ACAGTGACTTCCAAA..... 512
||| |||
798 CysValArgSerAspTyrThrGlyTyrValThrCysHisAsnSerAsnLe 814
513 .ACGAACAAGACACACCAGCAAAATTC...CCAGCCCTCTGTAGTTTATG 558
: : : : : ||| : : : : : ||| : : : : :
814 uSerGluLysAlaLeuAsnSerPheAsnProThrAsnLeuArgGlyAsnV 831
559 CAAATATTCCGCAAGGAGCCTCCCAATTC...AGGCCAGTCTCACA 605
: : : : : ||| : : : : : ||| : : : : :
831 aiAsnLeuThrGluAsnAlaSerPheThrLeuGlyLysAlaAsnLeuPhe 847
605 GCCCTGATTGAATCAGTGAATGGAAACACAGTTACCTTTGGAA..... 647
: : : : : ||| : : : : : ||| : : : : :
848 GlyThrIleGlnSerIleGlyThrSerGlnValAsnLeuLysGluAsnSe 864
648 .....CTACTGGATA 657
864 rHisTrpHisLeuThrGlyAsnSerAsnValAsnGlnLeuAsnLeuThra 881
658 ATGGA.....GCAGTGCTGATGCTACTAAGGATGACGGT 692
||||| ||| ||| : : : : : |||
881 snGlyHisIleHisLeuAsnAlaGlnAsnAspAlaAsnLys..... 894
693 GTCTACTCAAGGTATTTCAACTTATGACACGAATGGTAGATACAGTGT 742
||||| : : : : : ||| : : : : :
895 .....ValThrThrTyrAsnThr.....Le 901
743 AAAAGTGGGGCTCTGGGAGGAGTTAACCCAGCCAGGAGAGTGATAC 792
901 uThrValAsnSerLeuSerGly..... 908
793 CCCAGCAGAGTGGAGACTGTACATACCTGGCTGGATTGAGAATGATCAA 842
||| : : : : : ||| : : : : : ||| : : : : :
909 .....AsnGlySerPheTyrTyr.....TrpValAspPheThrAsn 920
843 ATACAATGGAATCCACCAAGACCTGAAATTAATAGGATGATGTTCAACA 892
: : : : : ||| : : : : : ||| : : : : :
921 AsnLysSerAsn.....LysValValValAsnLys..... 930
893 CAAGCAAGTGTGTTTCACGACAACATCCTCGGAGGCTCATTTGTGGCTT 942
||| : : : : : ||| : : : : : ||| : : : : :
943 CTGATGTCCCAATGCTCCCATACCTGATCTCTCCACCTGGCCAAATC 992
939 .....GlnVal 940
993 ACCGACCTGAAGCGGAAATTCACGGGGCAGTCTCAATT..... 1031
: : : : : ||| : : : : : ||| : : : : :
941 AlaAspLysThrGlyGluProAsnHisAsnGluLeuThrLeuPheAspAl 957
1032 .....AATCTCACTTGGACAGCTCCTCGGGGATGATT 1062
||| : : : : : ||| : : : : : ||| : : : : :
957 aSerAsnAlaThrArgAsnAsnLeuGluValThrLeuAlaAsnGlySerV 974
1063 ATGACCATGGAAACAGCTCACAAGTATATCATTCGAATAAGTACAGTATT 1112
||||| ||| ||| : : : : : ||| : : : : :
974 alaSpArgGly...AlaTrpLysTyrLysLeuArgAsnValAsnGlyArg 989
1113 CTTGATCTCAGA.....GACAAGTTCAATGAATCTCTTCAAGT 1150
: : : : : ||| : : : : : ||| : : : : :
990 TyrAspLeuTyrAsnProGluValGluLysArgAsnGlnThr.....Va 1004
1151 GAATFACTACTGCTCTCATCCCAAAGGAAGCAACTCTGAGGAAGTCTTTT 1200
||| : : : : : ||| : : : : : ||| : : : : :
1004 lAspThrThr..... 1007
1201 TGTTTAAACACAGAAACATTACTTTTGAATGGCACAGATCTTTTTCATT 1250
||||| |||

```


827 GATTGAGAAATGATGAATACAAATGGAATCCACCAAGACCTGAAATTAATA 876
446ArgLeuSerTrpArgProAlaGluAla.....L 456
877 AGGATGATGTTTCAACACACGAAGTGTCTTTCAGCAGA..... 914
456 yGlyAsnIleGlnThrPheThrValPhePheSerArgGluGlyAspAsn 472
915ACATCTCGGGAGGCTCATTT.....GT 937
473 ArgGluArgAlaLeuAsnThrThrGlnProGlySerLeuGlnLeuThrVa 489
938 GGCCTCTGATCTCCCAATGCT..... 959
489 lGlyAsnLeuLysProGluAlaMetTyrThrPheArgValValAlaTyrA 506
960CCATA..... 965
506 snGluTrpGlyProGlyGluSerSerGlnProIleLysValAlaThrCln 522
966 CCTGATCTCTCCACCTGCGCAATACACGACCTGAAGCGGAAATCA 1015
523 ProGluLeuGlnValProGlyProValGluAsnLeuGlnAlaValSerTh 539
1016 CGGGGGCAGTCTCATTATCTGACCTTGACAGACCTCCGCGGATGATATG 1065
539 rSerProThrSerIleLeuIleThrTrpGluProAla.....TyrA 554
1066 ACCATGGAACAGCTCACAAATATATCAAT.....CGAATAAGTACA 1106
554 laAsnGlyProValGlnGlyTyrArgLeuPheCysThrGluValSerThr 570
1107 AGTATCTTGATCTCAGACACAAGTTCAATGAATCTCTCAAGTGAATAC 1156
571 Gly..... 571
1157 TACTGCTCTCATCCCAAGGAGCAACTCTGAG...GAAGCTTTTGT 1203
572LysGluGlnAsnIleGluValAspGlyLeuSerT 583
1204 TTAACACAGAAACATTAATCTTTGAAATGGCACAGATCTTTTCATTCCT 1253
583 yrLysLeuGluGlyLeu.....LysLysPheThrGluTyrSerLeuArg 597
1254 ATTCAGGCTGTGTGATAAGTCTGATCTGAATCAGAAATATATCAACATTC 1303
598 PheLeuAlaTyrAsnArgTyrGlyProGlyValSerThrAspIleTh 614
1304 ACGAGTATCTTTGTTTATCTCCACAGACTCCGCCAGACACCTAGTC 1353
614 rValThrLeuSerAspValProSerAlaProGlnProGlnAsnValSerL 631
1354 CTGATGAACCTCTGCT.....CCTTGTCCT 1379
631 euGluValValAsnSerArgSerIleLysValSerTrpLeuProPro 647
1380 AATATTCATATCAACAGACCATTCCTGGCATTCACATT 1418
648 SerGlyThrGlnAsnGlyPheIleThrGlyTyrLysIle 660

seq_name: /cgn2_5/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US94-05277-2
seq_documentation_block:
; Sequence 2, Application PC/TUS9405277
; GENERAL INFORMATION:
; APPLICANT: Bruskin, Arthur
; APPLICANT: Jarosz, David E.
; APPLICANT: Johnson, Karen
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: Vogelstein, Bert
; APPLICANT: Zabrecky, James R.
; TITLE OF INVENTION: Antibodies Specific for DCC Gene Product

NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/05277
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 01107.42709
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202.508.9100
TELEFAX: 202.508.9299
TELEX: 197430 BBMB UT
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1447 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-05277-2

alignment_scores:
Quality: 115.00 Length: 563
Ratio: 0.475 Gaps: 26
Percent Similarity: 42.984 Percent Identity: 20.426

alignment_block:
US-09-049-696-19 x PCT-US94-05277-2 ..
Align:seg 1/1 to: PCT-US94-05277-2 from: 1 to: 1447
27 GTCCTTTGGGCGCCTCTGCAGCTCAAGACTAGAGAGCTGTCCAAAT 76
169 ValValValLeuProSerGlyAlaLeuGlnIleSerArgLeu.....G1 203
77 GACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGAACAAATGGCC 126
203 nProGlyAspIleGlyIleTyrArg..... 211
127 TCATTGATGCTTTTGGGCGCCTTTTCATCAGGAAATGGAGCTGTCTCTCAG 176
212 ..CysSerAlaArgAsnProAlaSerSerArgThrGlyAsnGluAlaGlu 227
177 CGCTCCATCCAGCTTGAGAGTAAGGATTA.....ACCTCCAGAACAG 220
228 ValArgIle...LeuSerAspProGlyLeuHisArgGlnLeuTyrPheLe 243
221 CCAGTGTGATGAATGGCACAGTATGCTGGCAGACACCGCTGGGAAGGACA 270
243 uGlnArgProSerAsnValValAlaIleGlu.....GlyLysAspA 257
271 CTTTGTGTTCTTATCACCTGGACACGCGACCTCCCAAATCCTTCTCGG 320
259 laValLeuGluCysValSerGlyTyrProProSerPheThrTrp 273
321 GATCCAGTGGACAGACAGAGTGGCTTTGTAGTGGACAAAAACACCAA 370
274LeuArgGlyGluGluValIleGlnLeuArgSerLy 285
371 AATGGCTACCTCCAAATCCAGGCAATGCTAAGTTGGCAGCTTGGAAAT 420


```
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-235-838-16

alignment_scores:
  Quality: 111.00      Length: 260
  Ratio: 0.881        Gaps: 12
  Percent Similarity: 48.462  Percent Identity: 20.769

alignment_block:
US-09-049-696-19 x US-08-235-838-16
..
Align seg 1/1 to: US-08-235-838-16 from: 1 to: 637

183 ATCCAGCTTGAGAGTAAGGATTACCTCCAGACACCCAGTGGATGAA 232
34 ValGlnLeuGlnGlnSerGlyAlaGluLeuValArgPro..... 46
233 TGGCAGACGATGATCGTGAGCAGCAGCGTGGAAAGGAC.....ACTT 273
47 .GlyThrSerVallyLeuSerCysLysAlaSerAspTyrThrPheThrS 63
274 TGTTCCTTATCACCCTGGACACGCGAG...CCTCCCAAAATCCTCTCGG 320
63 erTyrTrpMetAsnTrpVallyGlnArgProGlyGlnGlyLeuGluTrp 79
321 .....GATCCAGTGGACAGAG..... 338
80 IleglyMetIleAspProSerAspSerGluThrGlnTyrAsnGlnMetPh 96
339 .....CAAGGTGGCTTTGTAGTGGACAAACACCAAAATGGCTACC 381
96 eLysAspLysAlaAlaLeuThrValAspLysSerSerAsnThrAlaTyrM 113
382 TCCAAATCCAGGCAATCCTAAG..... 404
113 etGlnLeuSerSerLeuThrSerGluAspSerAlaValTyrTyrCysAla 129
405 .....GTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACA 442
130 LysGlyGlyAlaSerGlyAspTrpTyrPheAspValTrpGlyGlnGlyTh 146
443 AACCTTGACC..... 452
146 rThrValThrValSerSerGlyGlyGlySerGlyGlyGlySerG 163
453 .....CTGACTGTACGTCCTCCGTCGCTGCCAAT 479
163 lyGlyGlyGlySerAspIleGlnLeuThrGlnSerProSerSerLeuSer 179
480 GCTACCCCTG...CCTCCAATTACAGTACTTCCAAAGCAAGCAAGCAC 526
180 AlaSerLeuGlyGlyValThrIleThrCysLysAlaSerGlnAspIle 196
527 CAGCAAAATTCGCCAGCCCTCTGGTAGTTTATGCAAAATATTCGCCAAGGAG 576
196 eLysLysTyr.....IleAlaTrpTyrGlnHisLysProGlyL 209
577 CCTCCCAAAATTCAGGCCAGTGTACAGCCCTGATGAATCAGTGAAT 626
209 ysSerProArgLeuLeuIleHisTyrThrSerValLeuGln..... 222
627 GGAACAAACAGTTACCTTGGAACTACTGGATAATGGACAGGTGCTGATGC 676
```

```
223 ...ProGlyIleProSerArgPheSerGlySerGlySerGlyArgAspTy 238
677 TACT.....AAGGATGACGGTGTCTCTAC..... 698
238 rSerPheSerIleHisAsnLeuGluProGluAspIleAlaThrTyrTyrC 255
599 ..TCAAGGTATTTACAACTTATACACAGGAATGATAGATACAGTGTAAAA 746
255 ysLeuHisTyrAspTyrLeuTyrThrPheGlyGlyGlyThrLysLeuGlu 271
747 GTGCGGGCTCTGGCAGGAGTAAACGCACCC 776
272 IleLysAlaLeuGluGlySerLeuAla 281

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-465-473B-16
seq_documentation_block:
; Sequence 16, Application US/08465473B
; Patent No. 5939531
; GENERAL INFORMATION:
; APPLICANT: Wells, Winfried S.
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5939531man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NOVARTIS Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901-6940
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,473B
; FILING DATE: 5 June 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pfeiffer, Hesna J.
; REGISTRATION NUMBER: 22,640
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)522 6940
; TELEFAX: (908)522 6955
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-465-473B-16
```

```
alignment_scores:
  Quality: 111.00      Length: 260
  Ratio: 0.881        Gaps: 12
  Percent Similarity: 48.462  Percent Identity: 20.769
```

alignment_block:

US-09-049-696-19 x US-08-465-473B-16

Align seg 1/1 to: US-08-465-473B-16 from: 1 to: 637

```
183 ATCCAGCTTGAGAGTAGGATTAAACCTCCAGAACAGCCAGCTGGATGAA 232
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 ValGlnLeuGlnGlnSerGlyAlaGluLeuValArgPro..... 46
233 TGGCAGAGTAGCTGGACAGACCGCTGGGAAGGAC.....ACTT 273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
47 GlyThrSerValLysLeuSerCysLysAlaSerAspTyrThrPheThrS 63
274 TGTTCCTTATCCTCGGACGAACAG...CCTCCCAAAATCCTCTCTCGG 320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
63 eryTrpMetAsnTrpValLysGlnArgProGlyGlnGlyLeuGluTrp 79
321 .....GATCCAGTGGACAGAAG..... 338
80 IleGlyMetIleAspProSerAspSerGluThrGlnTyrAsnGlnMetPh 96
339 .....CAAGGTGGCTTGTAGTGTGACAAACACCAAAATGGCTTACC 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
96 eLysAspLysAlaLeuThrValAspLysSerSerAsnThrAlaTyrM 113
382 TCCAAATCCCGAGGCTTGTCTTAAG..... 404
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
113 etGlnLeuSerSerLeuThrSerGluAspSerAlaValTyrCysAla 129
405 .....GTTGGCCTTGGAAATACAGTCTCAAGCAAGCTCACAC 442
130 LysGlyGlyAlaSerGlyAspTrpTyrPheAspValTrpGlyGlnGly 146
443 AACCTTCAGC..... 452
146 rThrValThrValSerSerGlyGlyGlyGlySerGlyGlyGlySerG 163
453 .....CTGACTCTCAGCTCCCGCTCGCTCCAT 479
163 LyGlyGlyGlySerAspIleGlnLeuThrGlnSerProSerSerLeuSer 179
480 GCTACCTCTG...CCTCCAATACAGTCACTCCAAACAGCAAGACAC 526
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
180 AlaSerLeuGlyGlyGluValIleThrCysLysAlaSerGlnAspI 196
527 CAGCAAAATCCCGAGGCTCTGGTGTATGCAAAATATTCGCCAAGGAG 576
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
196 eLysLysTyr.....IleAlaTrpTyrGlnHisLysProGlyL 209
577 CTTCCCAATCTCAGGCGCAGTCTACAGCCCTGATTTGAATCAGTGAAT 626
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
209 ySerProArgLeuLeuIleHisTyrThrSerValLeuGln..... 222
627 GGAACAAACAGTTACCTGGAACTACTGGATAATGAGCAGCGTGTGATGC 676
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
223 ...ProGlyIleProSerArgPheSerGlySerGlyArgAspTyr 238
677 TACT.....AAGGATGACGGTGTCTAC.... 698
238 rSerPheSerIleHisAsnLeuGluProGluAspIleAlaThrTyrTr 255
699 ...TCAAGCTATTTCACACTTATCAGCAGGAATGCTAGATCAGTGTAAA 746
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
255 ysLeuHisTyrAspTyrLeuTyrThrPheGlyGlyGlyThrLysLeuGlu 271
747 GTGCGGCTCTGGGAGGAGTAAAGCGACCC 776
272 IleLysAlaLeuGluGlyGlySerLeuAla 281
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-540-245A-15

seq_documentation_block:

; Sequence 15, Application US/09540245A

```
Patent No. 6270984
GENERAL INFORMATION:
APPLICANT: Goodman, Corey
APPLICANT: Kid, Thomas
APPLICANT: Brose, Katja
APPLICANT: Tessier-Lavigne, Marc
TITLE OF INVENTION: Modulating Robo: Ligand Interactions
FILE REFERENCE: B98-031-3
CURRENT APPLICATION NUMBER: US/09/540,245A
CURRENT FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 60/065,544
PRIOR FILING DATE: 1997-11-14
PRIOR APPLICATION NUMBER: 60/081,057
PRIOR FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 1395
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-540-245A-15
```

alignment_scores:

```
Quality: 111.00 Length: 472
Ratio: 0.483 Gaps: 21
Percent Similarity: 48.729 Percent Identity: 21.186
```

alignment_block:

US-09-049-696-19 x US-09-540-245A-15

Align seg 1/1 to: US-09-540-245A-15 from: 1 to: 1395

```
144 GCCCTTTCATCAGGAATGGAGCTGTCTCTCAGCGCTCCATCCAGCTGA 193
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
484 AlaValGlnAlaGlyAsn.ArgTyrSerIleIleGlnGlySerSerLeu 500
194 GAGTAAGGATTAAACC.....TCCAGAACACCC 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
500 rGValAspAspLeuGlnLeuSerAspSerGlyThrTyrThrCysThrAla 516
223 ACTGGATGAATGGCACAGTGTCTGTGGACAGCACCGTGGAAAGACACT 272
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
517 SerGlyGluArgGlyGluThrSerTrpAlaAla.....ThrLe 529
273 TTGTTTCTTATCACCCTGGACACGACGCTC.....CCCA 307
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
529 uThrValGluTyrProGlySerThrSerLeuHisArgAlaAlaAspPro 546
308 AATCTTCTCTGGGATCCCA.....GTGGACAGA 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
546 eThrTyrProAlaProGlyThrProLysValLeuAsnValSerArg 562
337 AGCAGGTGGCTTTGTAGTGACAAACACCAAAATGGCCTACCTCCAA 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
563 ThrSerIleSerLeuArgTrpAlaLysSerGlnGluLysProGlyAlaVa 579
387 ATCCAGGCAATGCTAAGGTGGCAGCTTGGAAATACAGTCTGCAAGCAAG 436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
579 lGly...ProIleIle.GlyTyrThrValGluTyrPheSerProAspLeu 594
437 CTCACA.....AACCTTGACCTGTGACTGTACGTCC 468
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
595 GlnThrGlyTrpIleValAlaAlaHisArgValGlyAspThrGlnValTh 611
469 GTGCGTCCCAATGCTACCTCCCTCCCAATTTACAGTCACTTCCAAACGAC 518
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
611 rIleSerGlyLeuThrProGlyThrSerTyrValPheLeuValArgAlaG 628
519 AAGGACACACGACAAATTCGCCAGCCCTCTGGTAGTT..... 554
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
628 lAsnThrGlnGlyIleSerValProSerGlyLeuSerAsnValIleLys 644
```

```

#####
894 yAsnIlePro 897

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-728-470-10
seq_documentation_block:
; Sequence 10, Application US/08728470
; Patent NO. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstesser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-10
      :?
alignment_scores:
      Quality: 109.00 Length: 510
      Ratio: 0.458 Gaps: 22
Percent Similarity: 46.667 Percent Identity: 20.392

alignment_block:
US-09-049-696-19 x US-08-728-470-10 ..
--Align seg 1/1 to: US-08-728-470-10 from: 1 to: 1529

15   APTATCCACAGTGCCTTTGGGCCCTCTGCAAGTAAGACTAGAGGA 64
    ||| :::::::::::::::::::: :::: |||| :::: ||| ::
177 ILeThrTySerIleAlaLaProGUlaAsnGluaLleAlaLeuGIyAs 193
        65 GCGTGTCCAAATAACAGAGAGGTTCACAGCATATGCTTCAGATCAAGTTC 114
          :::: |::::::::::: :::: ||::: :::::
193 pRIePhEAlALysGIyGLYAsnILeAsnValAArgAlaA...ThrlEA 209
```


338 eAlAsythrGlyGlyPhe.....ValGluThrSerGlyHisAspL 352
 448 TGACCCCTG..... 455
 352 euSerIleGlyAspAspValIleValAspAlaLysGluTrpLeuLeuAsp 368
 456ACTGTCACGTCGCGTCGCTCCATGCTAC 484
 369 ProAspAspValSerIleGluThrLeuThrSerGlyArgAsnAsnThrG1 385
 485 CTGCGCTCCAATTACAGTGACTTCCAAACAGACAGACACCACGAA. 533
 385 yGluAsnGlnGlyThrThrGlyAspGlyThrLysGluSerProLysG 402
 534TTCCCCAGCGCCTGCTGCTAGTTTATGCATAATATTCGCGCAAGGA 575
 402 LyAsnSerIleSerLysProThrLeuThrAsnSerThrLeuGluGln... 417
 575 GCCTCCCCAATTCTCAGCGCCAGTGTCACAGCCCTGATTGAATCAGTGAA 625
 418IleLeuArgArgGlySerTyrValAsnIleThrAlaAsnAs 431
 626 TGGAAAACACGATTACCTTGGAACTA...CTGGATAATTGGACAGGTGCTG 672
 431 nArgIleTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeuThrL 448
 673 ATGCTACTACGAGTACGGTGCTCTACTCAAGGTATTTCACACACTTATGAC 722
 448 euHisThrLysArgAspGlyValLysIleAsnGlyAspIleThrSerAsn 464
 723 ACGAATGGTAGATACAGTGTAAAGTCGGGCTCTGGGAGGAGTTAAACGC 772
 465 GluAsnGlyAsnLeuThrIleLysAlaGlySerTrpValAspValHis... 480
 773 AGCCAGACGGAGAGTGATACCCACAGAGATGGAGCACTGTACATACCTG 822
 481LysAsnIleThrLeuGlyThrGlyPheLeuAsnIle... 492
 823 GCTGGATTGCAATGATGAAATACAATGGAATCCACCAAGACCTGAA... 869
 493ValAlaGlyAspSerValAlaPheGluArgGluGlyAspLysAla 507
 873 874
 508 ArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrIleThrValAs 524
 875 TAAGATGATGTTTCAACACAG.....CAAGTGTGTTTTCAGCAGACAT 918
 524 nLysAspAspLysGlnPheArgPheAsnAsnValSerIleAsnGlyThrG 541
 919 CTGCGGAGGCTCATTTGCTGCTCTGATGCCCAATGCTCCCATACCT 968
 541 LysLysGlyLeuLysPheIleAlaAsnGln..... 550
 969 GATCTCTCCACCTGGCCAAATCACC.....GACCTGAAGCGCGGA 1009
 551AsnAsnPheThrHisLysPheAspGlyGluLeuAs 562
 1019 AATTCACGGGGCAGTCTCATTAATCTGACT..... 1040
 562 nIleSerGlyIleValThrIleAsnGlnThrThrLysLysAspValLysT 579
 1041TGGACAGCTCCTGGGGATGATTATGACCAT..... 1070
 579 yTrpAsnAlaSerLysAspSerTyrTrpAsnValSerSerLeuThrLeu 595
 1071 GGAACAGCTACAAGTAT.....ATCATTCGAATAAGTACAAG 1108
 596 AsnThrValGlnLysPheThrPheIleLysPheValAspSerGlySerAs 612
 1109 TATTCTTGATCTCAGACACAGTTCAATGAATCTCTTCAAGTGAATACTA 1158

612 nserGlnAspLeuArgSerSerArgSerPheAlaGlyValHisPheA 629
; Sequence 19, Application US/08840466A
1159 CTGCTCTCATCCCAAGGAA.....GCCAACTCTGAGGAA 1193
: : : : : ||| : : : : :
629 snGlyIleGlyGlyLysThrAsnPheAsnIleGlyAlaAsnAlaLysAla 645
1194 GTCTTTTGTAAACACAGAAACATTACT 1223
: : : : : ||| : : : : :
646 LeuPheLysLeuLysProAsnAlaAlaThr 655

seq_name: /cgn2.6/ptodata/2/1aa/6B_COMB.pep:US-08-840-466A-19

seq_documentation_block:
; Sequence 19, Application US/08840466A
; Patent No. 6261561
GENERAL INFORMATION:
; APPLICANT: Stewart, C. Neal
; McKee, Marian L.
; O'Brien, Allison D.
; Wachtel, Marian R.
TITLE OF INVENTION:
; By Administration Of Stimulating An Immune Response
; By Administration Of Host Organisms That Express Intimin
; Alone Or As A Fusion Protein With One Or More Other
; Antigens.

NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
; ADDRESSER: Finnegan, Henderson, Farabow, Garrett &
; Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA

ZIP: 20005-3315
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840,466A
; FILING DATE: 18-Apr-1997
; CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
; NAME: Boone, Laura L.
; REGISTRATION NUMBER: 43,505
; REFERENCE/DOCKET NUMBER: 04995.0029-00000
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 934 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-08-840-466A-19

alignment_scores:
Quality: 107.50 Length: 513
Ratio: 0.418 Gaps: 30
Percent Similarity: 50.097 Percent Identity: 22.027

alignment_block:
US-09-049-696-19 x US-08-840-466A-19

Align seg 1/1 to: US-08-840-466A-19 from: 1 to: 934

3 CAAAGTGTGTCATCATCACACAGTCCCTTGGGGCCCTCTGAGCTCA 52
||||| ||| ||| : : : : :
495 GlnGlyGlyGlnIleGlnHisSer.....GlySerGlnSerAlaG 508

53 AGAACTAGAGGAGCTG.....TCCAAATACAGAGAGGTTTACAGACAT 96
: : : : : ||| : : : : :
503 nAspPyrGlnAlaIleLeuProAlaTyrValGlnGlyGlySerAsnIleT 525
: : : : : ||| : : : : :
97 ATGCTTCAGATCAAGTTCAGAACATGCCTCATTTGCTTTGGGGCC 146
: : : : : ||| : : : : :
525 yFLys.....ValThAlaArgAlaTyr 532
: : : : : ||| : : : : :
147 CTTTCATCAGGAATGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAG 196
: : : : : ||| : : : : :
533 AspArgAsnGlyAsn.....SerSerAsnAsnValGlnLeuThrI 546
197 TAAAGGATTAACTCCAGAACAGCCAGTGTGATGAATGGCACATGTCG 246
: : : : : ||| : : : : :
546 e.....ThrValLeuSerAsnGlyGlnValValAspGlnValGlyValT 561
: : : : : ||| : : : : :
247 TGGACAGCAGCGTGGGAAAG.....GACACTTTTGTCTTCTT 281
: : : : : ||| : : : : :
561 hrAspPheThrAlaAspLysThrSerAlaLysAlaAspAsnAlaAspThr 577
282 ATCACCTGACGACCAACGCCCTCCCAAAATCCTCTCTGGGATCCAGTGG 331
: : : : : ||| : : : : :
578 IleThrTyrThrAla.....ThrVal 584
332 ACAGAGCAAGGTGCTTTGTAGTGACAAAACACCAAAATGGCTACC 381
: : : : : ||| : : : : :
584 lLysLysAsnGly.....ValAlaGlnAlaAsnValProValSerPheA 599
382 TCCAAATCCAGGCATCTGTAAGTTGGCCTTTGGAATATACAGTCTGCAA 431
: : : : : ||| : : : : :
599 snIleValSerGlyThrAlaThrLeuGlyAlaAsnSerAlaLysThrAsp 615
432 GCAAGCTCACAACCTTGACCTGACTGTCAGTCCCGTCCGCTCCCAATGC 481
: : : : : ||| : : : : :
615 AlaAsnGlyLysAla.....ThrValThrLeuLysSerSer..... 627
482 TACCTGCTCCCAATACAGTGACTTCCAAAACAGCAAGGACACACGCA 531
: : : : : ||| : : : : :
628 ThrProGlyGlnValValSerAlaLysThrAlaGluMetSerSerA 644
532 AATTCCTCCAGCCCTCTGCTGAGTTTATGCAATATTTCGCCAAGAGCCCTCC 581
: : : : : ||| : : : : :
644 lAlaLeuAsnAlaSerAlaValIlePhePheAspGlnThrLysAlaSerIle 660
582 CCAATTCCTCAGGCCAGTCTCAGCCCTGATTGAATCAGTCAATGGAAA 631
: : : : : ||| : : : : :
661 ThrGluIleLysAlaAspLysThrThrAlaVal.....AlaAsnGlyL 675
632 AACAGTT.....ACCTTGGAACTACTGGATAATGGACAGAGTGTG 672
: : : : : ||| : : : : :
675 sAspAlaIleLysTyrThrValLysValMetLysAsnGly.GlnProVal 691
673 ATGCTACTAAGNATGACGGTGTCTACT.....CAAGGTAT 707
: : : : : ||| : : : : :
692 AsnAsnGlnSerValThrPheSerThrAsnPheGlyMetPheAsnGlyL 708
708 TTCACAACTTATGACA.....CGAATGGTAGATACAG 739
: : : : : ||| : : : : :
702 sSerGlnThrGlnAlaThrThrGlyAsnAspGlyArgAlaThrIleThrL 725
740 TGTAAAGTGGGGCTCTGGGAGGAGTTAACGACGACGACGAGAGTGA 789
: : : : : ||| : : : : :
725 eu.ThrSerSerAlaGlyLysAlaThrValSerAlaThrValSerAs 741
790 TACCCCGACAGAGTGGGACACTGTACATACCTGGCTGGATTGAG..... 833
: : : : : ||| : : : : :
741 pGlyAlaGluValLysAlaThrGluValThrPhePheAsp.GluLeuLys 757
834.....AATGATGAATACAAATGGAATCCACCAAGACCTGAAATTA 874
: : : : : ||| : : : : :
758 IleAspAsnLysValAspIleIleGlyAsnAsnValArgGlyGluLeuPr 774


```

? FILING DATE: 01-APR-1996
? CLASSIFICATION: 424
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/302,832
? FILING DATE: 05-OCT-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US PCT/US93/02166
? FILING DATE: 16-MAR-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Berkstresser, Jerry W
? REGISTRATION NUMBER: 22,651
? REFERENCE/DOCKET NUMBER: 1038-557
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (703) 415-0810
? TELEFAX: (703) 415-0813
? INFORMATION FOR SEQ ID NO: 10:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1600 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
?
? US-08-617-697-10
?
alignment_scores:
? Quality: 107.00 Length: 510
? Ratio: 0.450 Gaps: 22
? Percent Similarity: 46.667 Percent Identity: 20.392
?
alignment_block:
? US-09-049-696-19 x US-08-617-697-10

```

```

205 AACCTCCAGAACAGCCAGTGGAATGGCAGCAGTGCATCGTG..... 248
    : : : : : | | | | | : : : : :
314 eSerAlaGlnAsnGlnAlaLysGlyGlyLysLeuMetIleThrGlyA 331
    : : : : :
248 ..... 248

331 spLysValThrLeuLysThrGlyAlaValIleAspLeuSerGlyLysGlu 347
    : : : : :
249 .....GACAGCACCCTGGGAAAGCAGCAC 271
    : : : : : | | | | | : : : : :
348 GlyGlyGluThrTyrLeuGlyGlyAspIleArgGlyGlyLysAsnG 364
    : : : : : | | | | |
272 TTTGTTCTTATCACCTGGACACGACGCCCTCCCAAATCCTCTCTGGG 321
    : : : : : | | | | |
364 yIleGlnLeuAlaLysLysThrThrLeuGlyLysGlySerThrIle...A 380
    : : : : :
322 ATCCCACTGGACAGAAGCAAGGTGGC..... 347
    : : : : : | | | | | : : : : :
380 snValSerGlyLysGluLysGlyClyArgAlaIleValTrpClyAspIle 396
    : : : : :
348 TTTGTAGTGGACAAAACACCAAAATGGCTTACCTCCAAATCCCCAGGCAT 397
    : : : : : | | | | |
397 AlaLeuIleAspGlnValIleAsp.....AlaGlnIlySerAspIle 410
    : : : : :

```

1103 TATCTTGTATCTCAGACAAAGTTCAATGAATCTCTTCAAGTGAATCTA 1158
|||||
684 nGlyClnAspLeuArgSerArgSerPheAlaGlyValHisPheA 701
1159 CTGCTCTCATCCCAAGAA.....GCCAACTCTGAGGAA 1193
701 sNGLyIleGlyLysThrAsnPheAsnIleGlyAlaAsnAlaLysAla 717
1194 GTCCTTTTGTAAACCAAGAAACATTTACT 1223
718 LeuPheLysLeuLysProAsnAlaAlaThr 727
seq_name: /cgn2_8/ptodata/2/iaa/5A_COMB.pep.us-08-571-758-2

seq_documentation_block:
Sequence 2, Application US/08571758
Patent No. 5700675
GENERAL INFORMATION:
APPLICANT: Rubin, Gerry M.
APPLICANT: Therrien, Marc
APPLICANT: Chang, Henry C.
APPLICANT: Karim, Felix D.
APPLICANT: Wasserman, David A.
TITLE OF INVENTION: A No. 5700675el Protein Kinase Required for Ras
TITLE OF INVENTION: Signal Transduction
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/571,758
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B96-010
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 966 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
US-08-571-758-2

alignment_scores:
Quality: 106.50 Length: 458
Ratio: 0.595 Gaps: 20
Percent Similarity: 39.083 Percent Identity: 20.961
alignment_block:
US-09-049-696-19.x US-08-571-758-2 ..
Align seq 1/1 to: US-08-571-758-2 from: 1 to: 966

39 CCCTCTCAGCTCAAGAACTAGAGAGCTCTCCAAAATGACAGGAGTTT 88
|||||
244 ProSerAlaThrGluGluLeuAsn.....SerThrGlnGlySerGlnLe 258

398 TGCTAAGTTGGCACTTGGMAATACAGTCTGCNAGCAAGCTCACAACT 447
|||||
410 eAlaLysThrGlyGlyPhe.....ValGluThrSerGlyHisAspL 424
448 TGACCTG..... 455
424 eUerileGlyAspAspValIleValAlaLysGluThrLeuLeuAsp 440
456ACTGTCAGTCCCGTGCCTCAATCTAC 484
441 ProAspAspValSerileGluThrLeuThrSerGlyArgAsnAsnThrG 457
485 CCTGCTCCATTTACAGTACTCCAAAGCAAGACACCAAGCAAA. 533
457 yGluAsnGlnGlyThrThrGlyAspGlyThrLysGluSerProLysG 474
534TTCCCGAGCCCTCTGTTATGCAATATTCGCCAAGGA 575
474 yAsnSerileSerLysProThrLeuThrAsnSerThrLeuGluGln... 489
576 GCCTCCCAATTCACGGCCAGTGTACAGCCCTGATTCAGTCAGTGAA 625
490IleLeuArgGlySerTyrValAsnIleThrAlaAsnAs 503
626 TGGAAACACAGTTACCTTGGAACTA...CTGGATAATGGAGCGTCTG 672
503 nArgIleThrValAsnSerSerileAsnLeuSerAsnGlySerLeuThrL 520
673 ATGCTACTAAGATGACGGTGTCTACTCAAGGTATTTCAACACTTATGAC 722
520 euHisThrLysArgAspGlyValLysIleAsnGlyAspIleThrSerAsn 536
723 ACGATGTCATACAGTGTAAAGTGGCGCTCTGGGAGGATTAACGC 772
537 GluAsnGlyAsnLeuThrIleLysAlaGlySerTrpValAspValHis... 552
773 AGCCAGACGAGAGTGATACCCAGCAGAGTGGAGCAGTGTACATACCTG 822
553LysAsnIleThrLeuGlyThrGlyPheLeuAsnIle... 564
823 GCTGATGAGTCAAGTCAATACAAATGGAATCCACCAAGACCTGAA... 869
565ValAlaGlyAspSerValAlaPheGluArgGluGlyAspLysAla 579
870ATTAA 874
580 ArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrIleThrValas 596
875 TAAGATGATGTTCAACACAG.....CAAGTGTGTTTCAGCAGACAT 918
596 nLysAspAspLysGlnPheArgPheAsnAsnValSerLeuAsnGlyThrG 613
919 CTTCCGGAGGCTCATTTCTGCTGTCTGATGCCCAATGCTCCCATACCT 968
613 yLysGlyLeuLysPheIleAlaAsnGln..... 622
969 GATCTCTTCCACCTGGCCAAATCACC.....GACCTGAAGCGCGA 1009
623AsnAsnPheThrHisLysPheAspGlyGluIleAs 634
1010 AATTCACGGGGCAGTCTCATTATCTGACT..... 1040
634 nIleSerGlyIleValThrIleAsnGlnThrThrLysLysAspValLysT 651
1041 ...TGGACAGCTCTGGGAGTATTATGACCAT..... 1070
651 yTrpAsnAlaSerLysAspSerTyrTrpAsnValSerSerLeuThrLeu 667
1071 GGAACAGCTCACAGTAT.....ATCATTCGAAATGATACAG 1108
668 AsnThrValGlnLysPheThrPheIleLysPheValAspSerGlySerAs 684

```
89 ACAGACATATGCTTCAGATCAAGTTCAACAATGGCCTCATTGATGCTT 138
|   |   |   |   |   |   |   |   |   |   |   |   |   |
258 uThrLeuThrLeuThrProSerProAsnSerProPheThrProSer 275
|   |   |   |   |   |   |   |   |   |   |   |   |   |
139 TTGGGGCCCTTTCATCAGAAATGGAGCT   ..... 167
|   |   |   |   |   |   |   |   |   |   |   |   |   |
275 erGlyLeuSerSerSerLeuAsnGlyThrProGlnArgSerArgGlyThr 291
|   |   |   |   |   |   |   |   |   |   |   |   |   |
168 .....GTCTCTCAGCGCTCCATCCA 187
|   |   |   |   |   |   |   |   |   |   |   |   |   |
292 ProProAlaAlaGlyHisGlnThrLeuLeuSerGlnSerHisValG1 308
|   |   |   |   |   |   |   |   |   |   |   |   |   |
188 CTTTGAGAGTAAGGATTAACCTT   .....CCAGAACAGCCAGTGG 227
|   |   |   |   |   |   |   |   |   |   |   |   |   |
308 nValAspGlyGlnLeuAlaAlaAsnArgLeuProThrAspProSerT 325
|   |   |   |   |   |   |   |   |   |   |   |   |   |
228 ATGAATGCCAGATGATCGTGACAG   ..... 253
|   |   |   |   |   |   |   |   |   |   |   |   |   |
325 hrAspSerHisSerSerThrSerSerAspIlePheValAspProAsnThr 341
|   |   |   |   |   |   |   |   |   |   |   |   |   |
253 ..... 253
342 AsnAlaSerSerGlyGlySerSerSerAsnValLeuMetValProCysSe 358
|   |   |   |   |   |   |   |   |   |   |   |   |   |
254 .....CACCGTGGGAAGGACACTTTGTTTCTTATCACCTGG 290
|   |   |   |   |   |   |   |   |   |   |   |   |   |
358 rProGlyValGlyHisValGlyMetGlyHisAlaIleLysHisArgPheT 375
|   |   |   |   |   |   |   |   |   |   |   |   |   |
291 ACAACGAGCCTCCCAAAATCCTTCTCGGATCCAGTGGACAGAGCA 340
|   |   |   |   |   |   |   |   |   |   |   |   |   |
375 hrLysAla   .....LeuGlyPheMetAlaThrCysThr 385
|   |   |   |   |   |   |   |   |   |   |   |   |   |
341 AGGTGGCTTTAGTGACAAACACAAAATGGCTTACCTCCAAATCC 390
|   |   |   |   |   |   |   |   |   |   |   |   |   |
386 .....LeuCysGlnLysGlnValPheHisArg   ..... 394
|   |   |   |   |   |   |   |   |   |   |   |   |   |
391 CAGGCATTGCTAAGTTGGCCTTGGAAATACAGTCTCAAGCAAGCTCA 440
|   |   |   |   |   |   |   |   |   |   |   |   |   |
395 .....Trp   .....MetLysCyst 399
|   |   |   |   |   |   |   |   |   |   |   |   |   |
441 CAACACCTTGACCTGACTGTGCACCTCCGTCGCTCAATGCTACCTGCC 490
|   |   |   |   |   |   |   |   |   |   |   |   |   |
399 hrAspCysLysTyrlleCysHisLysSerCysAlaProHisValProPro 415
|   |   |   |   |   |   |   |   |   |   |   |   |   |
491 TCC   .....AATTACAGTGACTTCCAAA 513
|   |   |   |   |   |   |   |   |   |   |   |   |   |
416 SerCysGlyLeuProArgGluTyrrValAspGluPheArgHisIleLysG1 432
|   |   |   |   |   |   |   |   |   |   |   |   |   |
514 CGAACAGGACACAGCAAAATTCGCCAGCCCTCTGGTAGTTTATGCAAT 563
|   |   |   |   |   |   |   |   |   |   |   |   |   |
432 uGlnGlyGlyTyrrAlaSerLeuProHisValHisGlyAla   .AlaLysG 448
|   |   |   |   |   |   |   |   |   |   |   |   |   |
564 ATTCGCCAAGGAGCTCCCAATTCAGGCCAGTGTGTACAGCCCTGAT 613
|   |   |   |   |   |   |   |   |   |   |   |   |   |
448 lySerProLeuVal   ..... 452
|   |   |   |   |   |   |   |   |   |   |   |   |   |
614 TGAATCAGTGAATGGAACACAGTTACTTGGAACTACTGTGATAATGGAG 663
|   |   |   |   |   |   |   |   |   |   |   |   |   |
453 .....LysLysSerThrLeuGly   ..... 458
|   |   |   |   |   |   |   |   |   |   |   |   |   |
564 CAGGTGTGATGTACTAAGGATCAGCGTGTCTACTCAAGGTATTTCA 713
|   |   |   |   |   |   |   |   |   |   |   |   |   |
459 .....LysProLeuHis. 462
|   |   |   |   |   |   |   |   |   |   |   |   |   |
714 ACTTATGACAGCAATGGTAGATACAGTGTAAAGTGGCGCTCTGGGAGG 763
|   |   |   |   |   |   |   |   |   |   |   |   |   |
463 .....GlnGlnHisGlyAspSerSer 469
|   |   |   |   |   |   |   |   |   |   |   |   |   |
764 AGTTAAGCAGCCAGAGGAGTGTATACCCAGACAGTGGAGCATGT 813
|   |   |   |   |   |   |   |   |   |   |   |   |   |
470 SerProSerSerSerCysThrSerSerThrProSerSerProAlaLeup 486
|   |   |   |   |   |   |   |   |   |   |   |   |   |
814 ACATACCTGGCTGTGATGAGATGATGAATACAAATGGAATCCA   ..... 857
```

```
485 heGlnGlnArgGluArgGluLeuAspGlnAlaGlySerSerSerAla 502
|   |   |   |   |   |   |   |   |   |   |   |   |   |
858 .....CCAAGCCTGAAATTAATAAGCATGATGTTCAACACACGA 898
|   |   |   |   |   |   |   |   |   |   |   |   |   |
503 AsnLeuLeuProThrProSerLeuGlyLysHis   .....GlnProSerG1 517
|   |   |   |   |   |   |   |   |   |   |   |   |   |
899 AGTGTGTTTACGACA   .....ACATCCTCGGAGGCTCATTTGGG 939
|   |   |   |   |   |   |   |   |   |   |   |   |   |
517 nPheAsnPheProAsnValThrValThrSerSerGlyGlySerGlyGlyV 534
|   |   |   |   |   |   |   |   |   |   |   |   |   |
940 CTTCTGATGTCCTCCAAATGCTCCATACCTGATCTCTTCCCACCT   ..... 983
|   |   |   |   |   |   |   |   |   |   |   |   |   |
534 alSerLeuIleSerAsnGluProValProGluGlnPheProThrAlaPro 550
|   |   |   |   |   |   |   |   |   |   |   |   |   |
984 .....GCCAAATCACCAGCTGAAGCGGAAATTCACGGG   ..... 1019
|   |   |   |   |   |   |   |   |   |   |   |   |   |
551 AlaThrAlaAsnGlyGlyLeuAspSerLeuValSerSerAsnGlyHi 567
|   |   |   |   |   |   |   |   |   |   |   |   |   |
1019 ..... 1019
567 sMetSerSerLeuIleGlySerGlnThrSerAsnAlaSerThrAlaAlaT 584
|   |   |   |   |   |   |   |   |   |   |   |   |   |
1020 .....GCAGTCTCATTAATCTGACTTGGACAGCTCCTGGGATGAT 1061
|   |   |   |   |   |   |   |   |   |   |   |   |   |
584 hrLeuThrGlySerLeuValAsnSerThrThrThr   ..... 596
|   |   |   |   |   |   |   |   |   |   |   |   |   |
1062 TATGACCATGGAACAGCTCACAAAGTATATCATCGAATAAGTACAAGTAT 1111
|   |   |   |   |   |   |   |   |   |   |   |   |   |
597 .....SerThrCysSerPhePheProArgLysLeuSerThrAlaG1 610
|   |   |   |   |   |   |   |   |   |   |   |   |   |
1112 TCTTGATCTCAGACACAAGTTC 1133
|   |   |   |   |   |   |   |   |   |   |   |   |   |
610 yValAspLysArgThrProPhe 617
|   |   |   |   |   |   |   |   |   |   |   |   |   |
seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.us-08-909-984A-2
seq_documentation_block:
; Sequence 2, Application US/08909984A
; Patent No. 5747275
; GENERAL INFORMATION:
; APPLICANT: Rubini, Gerry M.
; APPLICANT: Therrien, Marc
; APPLICANT: Chang, Henry C.
; APPLICANT: Karim, Felix D.
; APPLICANT: Wasserman, David A.
; TITLE OF INVENTION: A No. 5747275el Protein Kinase Required for Ras
; TITLE OF INVENTION: Signal Transduction
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/909,984A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B96-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
```

SEQUENCE CHARACTERISTICS:
LENGTH: 966 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
US-08-909-984A-2

alignment_scores:
Quality: 106.50 Length: 458
Ratio: 0.595 Gaps: 20
Percent Similarity: 39.083 Percent Identity: 20.961

alignment_block:

US-09-049-696-19 x US-08-909-984A-2 ..

Align seg 1/1 to: US-08-909-984A-2 from: 1 to: 966

```
39 CCTCTGCAGCTCAAGACTAGAGGAGCTGTCCAAAATACACGAGGTTT 88
|||||
244 ProSerAlaThrGluGluLeuasn.....SerThrGlnGlySerGlnLe 258
|
89 ACAGACATATGCTTCAGATCAAGTTCAGAACAAATGCCCTCATTTGCTT 138
|
258 uThrLeuThrLeuThrProSerProAsnSerProPheThrProSerS 275
|
139 TTGGGGCCCTTCATCAGGAATGGAGCT..... 167
|
275 erGlyLeuSerSerSerLeuAsnGlyThrProGlnArgSerArgGlyThr 291
|
168 .....GTCTCTCAGCGCTCCATCCA 187
|
292 ProProProAlaArgLysHisGlnThrLeuLeuSerGlnSerHisValG 308
|
188 GCTTGAGAGTAAGGATTACCT.....CCAGAACCCAGTGG 227
|
308 nValAspGlyGluGlnLeuAlaArgAsnArgLeuProThrAspProSer 325
|
228 ATGATGGCAGCTGATCGTGGACAG..... 253
|
325 hrAspSerHisSerSerThrSerSerAspIlePheValaspProAsnThr 341
|
253 ..... 253
342 AsnAlaSerSerGlySerSerSerSerAsnValLeuMetValProCys 358
|
254 .....CACCGTGGAAAGGACACTTTGTTTCTTATCACCTGG 290
|
358 rProGlyValGlyHisValGlyMetGlyHisAlaIleLysHisArgPhe 375
|
291 ACAACGCGACCTCCCAATCCTTCTCTGGGATCCCGTGGACAGAACGA 340
|
375 hrLysAla.....LeuGlyPheMetAlaThrCysThr 385
|
341 AGGTGGCTTTCTAGTGACAAAAACACAAATGGCCTACCTCCAAATCC 390
|
386 .....LeuCysGlnLysGlnValPheHisArg..... 394
|
391 CAGGCATTGCTAAGGTGGCATTGGAAATACAGCTGCAGCAAGCTCA 440
|
395 .....TTP.....MetLysCys 399
|
441 CAACCTTGACCTGATGTCAGCTCCGGTCCGATGCTACCTGCC 490
|
399 hrAspCysLysTyrlleCysHisLysSerCysAlaProHisValProPro 415
|
491 TCC.....AATTACACTGACTTCCAAAA 513
|
416 SerCysGlyLeuProArgGluTyValAspGluPheArgHisIleLysG 432
|
514 CGAACAGGACACCAAGCAAAATCCCGACCTCTGGTAGTTTATGCAAA 563
|
```

```
432 uGlnGlyGlyTyTyAlaSerLeuProHisValHisGlyAla....AlaLysG 448
564 ATTCGCCAAGGAGCGCTCCCAATTCTCAGGGCCAGTGTACAGCCCTGAT 613
|||||
448 lySerProLeuVal..... 452
|
614 TGAATCAGTGAATGGAAAAACAGTTACCTTGGAACTACTGGATAATGGAG 663
|||||
453 .....LysLysSerThrLeuGly..... 458
|
664 CAGGTGCTGATGCTACTAAGGATGACGGTGTCTACTCAAGTATTTTCA 713
|
459 .....LysProLeuHis. 462
|
714 ACTTATGACACGAATGGTAGATACAGTGTAAAGTGGCGCTCTGGGAGG 763
|
463 .....GlnGlnHisGlyAspSerSer 469
|
764 AGTTAACGCCACCCAGAGCGGAGTGTATACCCACGACAGTGGACACTGT 813
|
470 SerProSerSerCysThrSerSerThrProSer.SerProAlaLeup 486
|
814 ACATACCTGGCTGGATTGAGAATGATGAAATACAAATGGAATCCA..... 857
|
486 heGlnGlnArgGluArgGluLeuAspGlnAlaGlySerSerSerAla 502
|
858 .....CCAAGACCTGAATAATAAGGATGATGTTCAACACACAGCA 898
|
503 AsnLeuLeuProThrProSerLeuGlyLysHis.....GlnProSerG 517
|
899 AGTGTGTTTTCAGCAGA.....ACATCCTCGGAGGCTCATTTGTTGG 939
|
517 nPheAsnPheProAsnValThrValThrSerSerGlySerGlyGlyV 534
|
940 CTTCTGATGTCCTCAAAATCTCCATACCTGATCTCTTCCACCT..... 983
|
534 alSerLeuIleSerAsnGluProValProGluGlnPheProThrAlaPro 550
|
984 .....GGCCAATCACCCACCTGAGGCGGAAATTCACGGG... 1019
|
553 AlaThrAlaAsnGlyGlyLeuAspSerLeuValSerSerSerAsnGlyH 567
|
1019 ..... 1019
567 smetSerSerLeuIleGlySerGlnThrSerAsnAlaSerThrAlaAla 584
|
1020 .....GGCAGTCTCATTAATCTGACTTGGACAGCTCCTGGGATGAT 1061
|
584 hrLeuThrGlySerLeuValAsnSerThrThr..... 596
|
1062 TATGACCATGGAAACAGCTCACAAGTATATATTCGAATAAGTACAAGTAT 1111
|
597 .....SerThrCysSerPhePheProArgLysLeuSerThrAlaGl 610
|
1112 TCTTGATCTCAGACAGCAAGTTC 1133
|
610 yValAspLysArgThrProPhe 617
|
seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-909-983-2
seq_documentation_block:
; Sequence 2, Application US/08909983
; Patent No. 5747288
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerry M.
; APPLICANT: Therrien, Marc
; APPLICANT: Chang, Henry C.
; APPLICANT: Karim, Felix D.
; APPLICANT: Wasserman, David A.
; TITLE OF INVENTION: A No. 5747288el Protein Kinase Required for Ras
; TITLE OF INVENTION: Signal Transduction
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
```

254CACGGTGGAAAGGACACTTGTCTTTATCATCCTGG 290
 350 rProGlyValGlyHisValGlyMetGlyHisAlaIleLysHisArgPhe 375
 291 ACAACGAGCGCTCCCAAAATCCTTCTGGATCCAGTGGACAGCAAGCA 340
 375 hrlYsAla.....LeuGlyPheMetAlaThrCysThr 395
 341 AGTGGCTTTGTAGTGACAAAACACCAAAATGGCGCTACCTCCAAATCC 390
 385LeuCysGlnLysGlnValPheHisArg..... 394
 392 CAGGCATTGCTAAGTGTGGCACTTGGAAATACACAGTCTGCAAGCAAGCTCA 440
 395Trp.....MetLysCyst 399
 441 CAACCTTGCACCTGACTGTACGTCCTCGCTCCGCTCCAAATGCTACCCTGCC 490
 399 hrAspCysLysTyrlIleCysHisLysSerCysAlaProHisValProPro 415
 491 TCC.....AATTACAGTCACCTCCAAAA 513
 416 SerCysGlyLeuProArgGluTyrlValAspGluPheArgHisIleLysG 432
 514 CGAACAAAGGACACAGCAAAATCCCCAGCGCCTCTGCTAGTATTTATGCAAA 563
 432 uGlnGlyGlyTyrlAlaSerLeuProHisValHisGlyAla...AlaLysG 448
 564 ATTTCGCAAGGAGCGCTCCCAATTCACAGGCGCAGTGTTCACAGCCCTGAT 613
 448 lySerProLeuVal..... 452
 614 TGAATCAGTCAATGGAAAAACAGTTTACCTTGGAACTACTCGATAATGGAG 663

453LysSyrserinLeuGly..... 458
664 CAGGTCCTGATGCTACTAAGGATGACGGTGCTACTCAAGTATTTCACA 713
459LysProLeuHis..... 462

714 ACTTATGACACGAATGGTAGATACAGTGTAAAAGTGGGGCTCTGGGAGG 763

463G:\G\HIS\G\Y\aspser 469

764 AGTTAAACGACGCCAGACGGAGAGTGATACCCAGACAGATGGAGCACTGT 813
 470 SerProSerSerSerCysThrSerSerThrProSer.SerProAlaLeuP 486

814 ACATACCTGGCTGGATTGAGAATGATGAATAACAATGGAAATCCA..... 857
:: ||| |||::: :::
486 heGlnGlnArgGluArgGluLeuAspGlnAlaGlySerSerSerAla 502

858CCAAGACCTGAAATTAAATAAGGATGATGTTCAACACACAAGCA 898
503 AsnLeuLeuProThrProSerLeuGlyLysHis.....GlnProSerG1 517

899 AGTGTGTTTCAGCAGA.....ACATCCTCGGAGGCTCATTTGTGG 939
 517 nPheAsnPheProAsnValThrValThrSerSerGlyGlySerGlyGlyVal 534

[illegible]

984GGCCAAATCACCGACCTGAGCGGAATTTCACGGG.. 1019
:: :: :: :: :: :: :: :: :: :: :: :: ::
551 AlaThrAlaAsnGlyGlyLeuAspSerLeuValSerSerAsnGlyH1 567

1019 1019
567 sMetSerSerLeuIleGlySerGlnThrSerAsnAlaSerThrAlaAlaT 584

11020GGCAGTCTCATTAATCTGACTTGGACAGCTCCTGGGATGAT 1061

```
|||||..... 584 hrLeuThrGlySerLeuValAsnSerThrThrThr..... 596
* 1062 TATGACCATGAACAGCTCACAGTATATATCATTCGAAATAGTACAGTAT 1111
:|||||: 597 .....SerThrCysSerPheProArgLysLeuSerThrAlaG1 610
1112 TCTTGATCTCAGACAGATTC 1133
:||||| 610 yValAspLysArgThrProPhe 617
```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-540-245A-18

```
seq_documentation_block:
; Sequence 18, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1651
; TYPE: PRT
; ORGANISM: human
US-09-540-245A-18
```

```
alignment_scores:
Quality: 106.00 Length: 466
Ratio: 0.493 Gaps: 24
Percent Similarity: 46.137 Percent Identity: 20.601
```

alignment_block:

US-09-049-696-19 x US-09-540-245A-18 ..

Align seg 1/1 to: US-09-540-245A-18 from: 1 to: 1651

```
246 GTGACACACCGCTGGGAAGGACACTTGTCTTCTATCACTGG..... 290
|||||..... 462 ValAsnGlnThrValAlaValAspGlyThrPheValLeuSerCysValAl 478
291 ..ACAAGCAGCGCTCCCAAAATCCCTCTCTGGGATCCCGAGTGGACAGAAGC 339
||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
478 aThrGlySerProValProThrIleLeuTrp.....ArgLysA 491
340 AAGTGGCTTTGTAGTGGACAAAACCAAAATGGCTACCTCCAATC 389
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
491 spGlyValLeuValSerThrGlnAspSerArgIleLysGlnLeuGlu... 506
390 CCAGGCATT.....GCTAAGTTGGC...ACTTGGAAATA 421
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
507 AsnGlyValLeuGlnIleArgTyAlaLysLeuGlyAspThrGlyArgTy 523
422 CAGTCTGCAGCAAGCTCACAAACCTTGACCTGACTGTACAGTCC.... 467
||||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..||| ..|||
523 rThrCysIleAlaSerThrProSerGlyGluAlaThrTrpSerAlaTyri 540
468 .....CGTGGCTCCAAT 479
540 leGluValGlnGluPheGlyValProValGlnProProArgProThrAsp 556
480 GCTACCTGCCCTCCAATTACAGTACTGCTTCCAAAACGAACAGGACACCA 529
```

```
..... 557 ProAsnLeuIleProSerAlaProSerLysProGluValThrAspValSe 573
||||| ||| ...: 530. CAATATCCCGAGCGCTCGGTAGTTTATGCAATATATTCGCAAGGAGCCT 579
||||| 573 rArgAsnThrValThrLeuSerTrpGlnProAsnLeuAsnSerGlyAlaTr 590
580 CCCCC.....ATTCTCAGGGCC.....AGT 599
:||||| 590 hrProThrSerTyriIleIleGluAlaPheSerHisAlaSerGlySerSer 606
:||||| 600 GTCACAGCGCTGATTTGAATCACTAGTGAATCGAAACAAACAGTTACCTTGGAA... 647
:||||| 607 TrpGlnThrValAlaGluAsnValLysThrGluThrSerAlaIleLysG1 623
648 .....CTACTGGATAAATGGACGAGTGTCTG 672
623 yLeuLysProAsnAlaIleTyLeuPheLeuValArgAlaAlaAsnAlaTr 640
673 ATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTTACACACTTATGAC 722
:||||| 640 yArgLysSerAspProSerGlnIleSerAspProValLysThrGlnAsp 656
723 .....ACCAATGGTAGATACAGTGTAAAGTG...CGGGCTCT 757
657 ValLeuProThrSerGlnGlyValAspHisLysGlnValGlnArgGluLe 673
759 GGGAGGAGTTAACGCGACGACGACGAGGA..... 785
||||| 673 uGly.....AsnAlaValLeuHisLeuHisAsnProThrValLeuSerS 688
786 .....GTGATACCCACGACAGT..... 803
683 erSerSerIleGluValHisTrpThrValAspGlnGlnSerGlnTyriLe 704
804 .....GGAGCACCTGTACATACCTGGCTGGATTGAGAAATGATGAAT 844
||||| 705 GlnGlyTyriLysIleLeuTyriArgProSerGlyAlaAsnHisGlyGluSe 721
845 ACAATGG.....AATCCA..... 857
721 rAspTrpLeuValPheGluValArgThrProAlaLysAsnSerValValI 738
858 .....CCAAAGACCTGAA 869
738 leProAspLeuArgLysGlyValAsnTyriGluIleLysAlaArgProPhe 754
870 ATTAATAAGGATGATGTTCAACACAGCAAGTGTGTTTTCAGCAGACATC 919
:||||| 755 PheAsnGluPheGlnGlyAlaAspSerGluIleLysPheAlaLysThr... 770
920 CTCGGGAGGCTCATTTGTGGCTCTGATGTCACAAATGCTCCATACCTG 969
771 .....LeuGluGluAlaProSerAla..... 777
970 ATCTCTTCCACCTGGCCAAATCACCGACCTGAAGCGCGAAATTCACGGG 1019
773 .....ProProGlnGlyValThrValSerLysAsnAspGlyAsnGly 791
1020 GGAGTCTCAATTAATCTGACTTGGCAGCTCCGCGGATGATATATGACCA 1069
792 ThrAlaIleLeu...ValSerTrpGlnProProGluAspThrGlnAs 807
1070 TGGACAGCTCACAGTATATCATTCGAATAAGTACAAGTATTCTTGTATC 1119
807 nGlyMetValGlnGluTyriLysValTrpCysLeuGlyAsn.....GluT 822
1120 TCAGAGACAAAGTTCAATGAATCTTCAAGTGAATACTACTGCT..... 1163
822 hrArgTyriHisIleAsnLysThrValAspGlySerThrPheSerValVal 838
1164 .....CTCATCCCAAGGAAGCACTCTGAGGAAGTC..... 1196
```

```

839 IleProPheLeuValProGlyIleArgTyrSerValGluValAlaAlaSe 855
1197 .....TTTTCTTTAAACACGAGAAACATTACTTTTG 1227
      ::|||::: ||| ::::
855 rThrGlyAlaGlySerGlyValLysSerGluProGlnPheIleGlnLeuA 872
1228 AA...AATGGCACAGATCTTTTCATTGCTATTTCAGGCTGTTGATAAGGTC 1274
      ::::|||::: ::::|||::: |||::|||
872 spAlaHisGlyAsn.....ProValSerProGluAspGlnVal 884
1275 GATCTGAATCAGAAATATCCAACATTCACGAGTATCTTTGTTTATT 1322
      ::||| ::::|||::|::|::|::|::|::|::|::|::|::|::|
885 SerLeuAlaGlnGlnIleSerAspValValLysGlnProAlaPheIle 900

```

This Page Blank (uspto)

552 nPheileThrAsnLeuThrPheArgThrAlaSerLeuTrpIleProGlyT 569
397 TTGCTAAGGTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAAACC 446
569 hrAlaLysProGlyHisTrpThrTyrThrLeuMetCysPheHisAla 585
447 TTGACCTGACT 458
586 LysLeuLeuThr 589

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA11324

seq_documentation_block:
ID AAB11324 standard; Protein; 592 AA.
AC AAB11324;
DT 21-FEB-2001 (first entry)
XX Human lung cancer-associated protein L762P variant 1.
DE Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW vaccine; detection.
XX Homo sapiens.
OS WO200061612-A2.
PN 19-OCT-2000.
PD 03-APR-2000; 2000WO-US08896.
XX 02-APR-1999; 99US-0285479.
PR 17-DEC-1999; 99US-0466396.
PR 30-DEC-1999; 99US-0476496.
PR 10-JAN-2000; 2000US-0480884.
PR 22-FEB-2000; 2000US-0510376.
XX (CORI-) CORIXA CORP.
XX Wang T, Fan L;
XX WPI: 2000-628399/60.
DR N-PSDB; AAC65896.

Isolated polypeptide comprising an immunogenic portion of a lung tumor protein is used for detecting and monitoring progression of lung cancer in a patient -

Claim 3; Page 177-178; 261pp; English.

This invention describes a novel isolated polypeptide (I) which comprising an immunogenic portion of a lung tumor protein or variant (P2) CC. which have cytostatic activity. The polypeptides and polynucleotides are CC used in compositions and vaccines to inhibit the development of cancer, CC especially lung cancer, in a patient. Methods described in the invention CC can be used to monitor the progression of a cancer by carrying out the CC detection at subsequent time points and comparing the results from the CC different time points. CD4+ and/or CD8+ T-Cells isolated from a patient CC are treated with p2, polynucleotides encoding p2 or antigen presenting CC cells expressing p2, and then administered to the patient to inhibit CC development of cancer.

XX Sequence 592 AA;

alignment_scores:
Quality: 343.50 Length: 154
Ratio: 2.887 Gaps: 2
Percent Similarity: 77.273 Percent Identity: 46.104

alignment_block:

US-09-049-696-19 x AAB11324 ..
Align seg 1/1 to: AAB11324 from: 1 to: 592
6 AGTGTGTCATCATCCACACAGTCGCTTTGGGCGCCTCTGCAGCTCAAGA 55
436 SerGlySerThrIleHisSerIleAlaLeuGlySerSerAlaAlaProAs 452
56 ACTAGAGGAGCTGTCCAAATGACAGAGGAGTTTACAGACATATGCTTCAG 105
452 nLeuGluLeuSerArgLeuThrGlyGlyLeuLysPheValProA 469
106 ATCAAGTTTCAGAAACAAATGCCCTCATGTATGCTTTTGGGGCCCTTTCATCA 155
469 spIleSerAsnSerAsnMetIleAspAlaPheSerArgIleSerSer 485
156 GGAATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGGATT 205
486 GlyThrGlyAspIlePheGlnGlnHisIleGlnLeuGluSerThrGlyG1 502
205 AACCTCCAGAACACGCCAGTGGATGAATGGCACAGTATGCTGGACAGCA 255
502 uAsnValLysProHisHisGlnLeuLysAsnThrValThrValAspAsnT 519
256 CGTGGGAAGGACACTTGTCTTATCACCTGG...ACAACGACGCGCT 302
519 hrValGlyAsnAspThrMetPheLeuValThrPGLnAlaSerGlyPro 535
303 CCCCAATCTCTCTGGGATCCCGAGTGACAGAAG.....CAAGGTGG 346
536 ProGluIleIleLeuPheAspProAspGlyArgLysTyrTyrThrAsnAs 552
347 CTTGTAGTGGACAAACACCAACAAATGGCTACCTCCAAATCCAGGCA 396
552 nPheIleThrAsnLeuThrPheArgThrAlaSerLeuTrpIleProGlyT 569
397 TTGCTAAGGTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAAACC 446
569 hrAlaLysProGlyHisTrpThrTyrThrLeuMetCysPheHisAla 585
447 TTGACCTGACT 458
586 LysLeuLeuThr 589

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA11889

seq_documentation_block:
ID AAB31889 standard; Protein; 4393 AA.

XX AAB31889;
XX 15-MAY-2001 (first entry)
XX Amino acid sequence of a human protein.
XX Euman; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;
KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;
KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;
KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;
KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.
XX Homo sapiens.
XX WO200105422-A2.
XX 25-JAN-2001.
XX 17-JUL-2000; 2000WO-FR02057.
XX 15-JUL-1999; 99FR-0009372.
XX (INMR) BIOMERIEUX STELHYS.

.....


```
929 CTCATTTCTGCTCTGATGTCCTCCCAATGCTCCCATACCTCATCTCTTCC 978
223 .....AlaAspAlaThrGlyAlaValLeu..... 230
979 CACCTGGCCAAATCACCCACCTGAAGCGGAAATTCAC..... 1016
231 .....SerIleThrAspAlaLysGlyAsnLeuGlnArgValAlaTyr 244
1017 .....GGGGGCGAGT..... 1025
245 AspValAlaGlyLeuLeuSerGlySerTrpLeuThrLeuLysAspGlyTh 261
1026 .....CTCATTAATCTGACTTGGCAGCTCCTGGG..... 1055
261 rGluGlnValIleValAlaSerLeuThrTyrSerAlaAlaGlyLysLysL 278
1056 .....GATGATTATGACCATGCAAGCTCAAGTATATATTCGGAATA 1100
278 euArgGluGluHisGlyAsnGlyValValThrSerTyrIleTyrGluPro 294
1101 AGTCAAGTATTCCT..... 1115
295 GluThrGlnArgLeuThrGlyIleLysThrGluArgProSerGlyHisVa 311
1116 .....GATCTCAGA.....CACAGTTCA 1134
311 lAlaGlyAlaLysValLeuGlnAspLeuArgTyrThrTyrAspProValG 328
1135 ATGATCTCTTCAAGTGAATACTACTGCTCTCATCCCAAGGAAGCCAAAC 1184
328 lYAsnValLeuSerValAsnAsnAspAla..... 337
1185 TCTGAGGAAGTCTTTTCTTTAAACGAGAAACATTACTTTTGAATATGG 1234
338 ...GluGluThrArgPheTrpArgAsnGlnLysValProGluAsnTh 353
1235 C.....ACAGATCTTTTCATTCGCTATTCAGGCTCTTGTAAAGTCG 1275
353 rTyrIleTyrAspSerLeuTyrGlnLeuValSerAlaThrGlyArg.... 368
1276 ATCTGAATCATCAATATCCACAGATTCGACGAGTATCTTTGTTTATTCCT 1325
369 .....GluMetAlaAsnAlaGlyGlnGlnGlyAsnAspLeuPro 381
1326 CCACAGACTCGCCAGACACCTAGTCTCTGATGAAGCTGCT 1370
382 SerAlaThr.....AlaProLeuProThrAspSerSerAla 393
```

seq_name: /SIDS2/gcdata/geneseq/geneseq/AA2001.DAT.AAU05396

seq_documentation_block:

ID AAU05396 standard; Protein: 26926 AA.

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

XX AC AAU05396;

PA (GPHO) GEN HOSPITAL CORP.

PI Enshman MC;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

alignment_scores:

Quality: 121.00 Length: 517

Ratio: 0.545 Gaps: 29

Percent Similarity: 42.940 Percent Identity: 19.729

alignment_block:

US-09-049-696-19 x AAU05396

Align seg 1/1 to: AAU05396 from: 1 to: 26926

145 CCCTTTCATCAGGAAATGGAGCTGCTCTCAGCGCTCCATCCAGCTTGAG 194

7075 ProAlaTyrValAspGluProValAsnMetSerThrProAlaThr..... 7089

195 AGTAAGGATTACCTCCAGACAGCCAGTGGATGAATGCCACAGTGTAT 244

7690 ValProAspProPro..GluAsnValLysTrpArgAspArgThrAla.. 7104

245 CGTGGACAGCACCGTGGGAAAGGACACTTTGTTCTTATCACCCTGGACAA 294

7105AsnSerIlePheLeu...ThrTrp..... 7111

295 CGCAGCCTCCCAAAATCCTTCTCTGGGATCCAGTGGAGAGCAAGCAAGGT 344

7113 AspProProLys.....AsnAspGlyGlySerArgIleLys 7123

345 GGCTTTGTAGTGGACAAA..... 362

7124 GlyTyrIleValGluArgCysProArgGlySerAspLysTrpValAlaLacY 7140

363AACACCAAATGGCTACCTCCAAATCCAGGCA 396

7140 sGlyGluProValAlaGluThrLysMet.....GluValThrGlyL 7154

397 TTGCTAAGGTTGGCAGCTTGG.....AAATACAGCTCTG 428

7154 euGluGlu...GlyLysTrpTyrAlaTyrArgValLysThrLeuAsnArg 7169

```

1113 CTTGATCTCAGAGACAAGTTC.....AATGAATCTCTTCAAGTGAATAC 1155
      |||||:::  :::::  |||:::
7429 .....ThrAspLysTrpValArgCysAsnLysMetProValIysAs  7442
      |||||:::  :::::  |||:::
1157 TACTGCTCTCATCCCAAAGGAAGCAACTCTCAGGAAGTCTTTTCTTTTA 1206
      |||||:::  |||  :::::  |||:::  |||:::
7442 pThrThrTyrArgValIysGlyLeuThrAsnLysLysLysTyrArgPhea  7459
      :::  |||||  :::
1207 AACCAAGAAAACATCTACTTTTGAATATGGCAGACAGTCTTTTCATGCTATT 1256
      :::  |||||  :::
7459 rg.....ValLeuAlaGluAsnLeuAlaGlyProGlyLysProSer  7472
      :::  |||||  :::
1257 CAGCGCTGTGATAGGTCGATCTGAATCATCAGAAATATCCAACATTGCACG 1306
      :::  |||||  :::
7473 LysSerThrGluProIleLeuIleLysAspProIleAsp.....  7485
      :::  |||||  :::
1307 AGTATCTTTGTTTATTCTCCACAGACTCCGCCAGAGACACCTAGTCTCG 1356
      |||  |||||  |||:::
7486 .....ProProTrpProGlyLysProThrVal  7496
      |||||  |||||  |||:::
1357 ATGAACACTCTGCTCCTCTGTCCTCATATATTCATATCAACAGCACCATTCT 1406
      :::::  :::::  |||  |||  |||
7496 ysAspValGlyLys.....ThrSerValArgLeuAsnTrpThrLysPro  7510
      :::  |||||  :::
seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:AA195203
seq_documentation_block:
AA195203 standard; Protein: 400 AA.
AA195203:
10-MAY-1996 (first entry)
Tyrosine phosphatase MPTP-delta insertion sequence, INS-2.
Tyrosine phosphatase MPTP-delta; murine; brain tissue;
glutathione-S-transferase; fusion protein; E. coli; differ-
entiation; information transmission; nervous system; immu-
nogenesis; insertion; INS-1; INS-2.
Mus musculus domesticus.
JP07236487-A.
12-SEP-1995.
28-FEB-1994; 94JP-0054726.
28-FEB-1994; 94JP-0054726.
(TOKS-) TOKYOTO SHINKEI KAGAKU SOGO KENKYUSHO ZH.
WPI; 1995-347455/45.
N-PSDB; AAO94313.
DNA encoding tyrosine phosphatase MPTP delta - useful for
elucidation of signal transmission mechanisms.
Claim 2; Page 12-13; 14pp; Japanese.
The sequences given in AAR75202-03 are encoded by insertion
sequences which were included in the tyrosine phosphatase MPTP-delta
sequence. The INS-1 cDNA was inserted between the 66th A and the
G of the MPTP-delta coding sequence, and the INS-2 cDNA was
inserted between the 1194th A and the 1195th T of the sequence. The
sequence was isolated from murine brain tissue and was cloned into a
expression vector, into the downstream region of a glutathione-S-
transferase gene, and expressed as a fusion protein in E. coli. MPTP-
delta proteins regulate differentiation and activation of cells
sequence can be used in the elucidation of the molecular mechanism
information transmission in cells, regulation mechanisms in the
system or immune system, or in the mechanism of carcinogenesis.

```


698 CTTCAAGGTATTTCACAACTTATGACACAAGAATGGTAGTATACAGTCTAAAA...746
::: ::: |||::: |||::: |||:::
398 rGlncysValAlaGluAsnGluAlaGlyAsnAlaGlnThrSerAlaGlnL 415
747GTCCGGGCTCTGGGAGGACTTAACGCAGCC 776
::: ::: |||::: |||::: |||:::
415 eufileValProLysProAlaIleProSerSerValLeuProSerAla 431
777 AGACGCAGAGTGATACCCCAGCAGAGTGAGCACGTGTACATACCTGGCTG 826
||| |||:::||| :::: |||:::
432 ProArgSpValValProValLeuValSerSerArgPheVal..... 445
827 GATTGAGAATGATAAATCAATGAATCCACCACAGACCTGAATTAATA 876
::: |||::: |||::: |||::: |||:::
446ArgLeuSerTrpArgProProAlaGluAla.....L 456
877 AGGATGATCTTCAACAACAAGCAAGTGTGTTTCAGCAGA..... 914
||| :::: |||::: |||::: |||:::
456 ySgLyAsnIleGlnThrPheThrValPhePheSerArgGluGlyAspAsn 472
915ACATCCTCGGAGGCGTCATTT.....GT 937
::: |||::: |||::: |||::: |||:::
473 ArgGluArgAlaLeuAsnThrThrGlnProGlySerLeuGlnLeuThrVa 489
938 GGCTTCCTGATGCCCAAATGCT..... 959
||||::: |||::: |||::: |||:::
489 lGlyAsnLeuLysProGluAlaMetThrPheArgValValAlaTyra 506
960CCCAT..... 965
snGluTrpGlyProGlyGluSerSerGlnProIleLysValAlaThrGln 522
966 CCTGATCTCTCCACCTGCCAAAATCACCGACCTCAAGCGGGAATTCA 1015
|||:::|||:::|||:::|||:::|||:::|||:::|||:::
523 ProGluLeuGlnValProGlyProValGluAsnLeuGluAlaValSerTh 539
1016 CGGGGGCAGTCTCATTAATCTGACTTGGACAGCTCCTCGGGGATGATTNG 1065
::: :: ||| ::|||:: |||::: |||::: |||:::
539 rSerProThrSerIleLeuIleThrTrpGluProProAla.....Tyra 554
1066 ACCATGGAACAGCTCAAGAATATATCATTT.....CGAATAAGTACA 1106
laasnGlyProValGlnGlyTyArgLeuPheCysThrGluValSerThr 570
1107 AGTATCTTGATCTCAGACAGACAGTTCATGAATCTCTCCAAGTGAATC 1156
::: Gly..... 571
1157 TACTGCTCTCATCCAAAGGAAGCAACTCTGAG...GAAGTCTTTTTTGT 1203
|||:::|||:::|||:::|||:::|||:::|||:::|||:::
572LysGluGlnAsnIleGluValAspGlyLeuSer 583
1204 TTAAACACAGAAAACATPACTTTTGAANAATGCCACAGATCTTTCATGTCT 1253
::||| |||:::|||::: :::: |||::: |||:::
583 TyrLysLeuGluGlyLeu.....LysLysPheThrGluTySerLeuArg 597
1254 ATTCAAGGCTGTGTATAAGGTCGATCTGAATTCAGAAATATCCACATATGC 1303
::: |||::: |||::: |||::: |||::: |||:::
598 PheLeuAlaTyArgAsnArgTyrglyProGlyValSerThrAspAspIleTh 614
1304 ACCAGTATCTTTGTTTATTCCTCCACAGACTCCGCCAGACAGACCTAGTC 1353
:: |||:::||| ::|||:::|||:::|||:::|||:::|||:::
614 rValValThrLeuSerAspValProSerAlaProGlnAsnValSerL 631
1354 CTGATGAACGCTGCT.....CCTTGTCCT 1379
::: Gly.....
631 euGluValValAsnSerArgSerIleLysValSerTrpLeuProPro 647
1380 AATATTATATATCAACAGACACCAANTCTCGCATTCACATT 1418
:: |||::: |||::: |||::: |||:::
648 SerGlyThrGlnAsnGlyPheIleThrGlyTyrsIle 660

203 nProGlyAspIleGlyIleTyrArg.....

```

127 TCATTGATGCTTTTGGGGCCCTTTTCATCAGGAATGAGCTGCTCTCAG 176
212 ..CysSerAlaArgAsnProAlaSerArgThrGlyAsnGluAlaGlu 227
177 CGCTCCATCCAGCTTGAGAGTAAGGGATTA.....ACCCTCCAGAACAG 220
228 ValArgile...LeuSerAspProGlyLeuHisArgGlnLeuTyrPheLe 243
221 CCAGTGGATGAATGGACAGTATCGTGGACAGCACCGTGGGAAGACA 270
243 uGlnArgProSerAsnValValAlaIleGlu.....GlyLysAspA 257
271 CTTTGTGTTCTTATCACCTGGACACAGCGCTCCCAAAATCCTTCTG 320
257 laValLeuGluCysValSerGlyTyrProProSerPheThrTirp 273
321 GATCCAGTGGACAGAGAGAGTGGCTTTGTAGTGGACAAAACACCAA 370
274 .....LeuArgGlyGluGluValIleGlnLeuArgSerLy 285
371 ANTGGCTACTCTCAAAATCCAGGCATGTAGGTTGGCACTTGGAAAT 420
285 s.....LysT 287
421 ACAGTCTGCAAGCAAGCTCACAACCTTGACCCCTGACTGTCACGTC 470
287 yrSerLeuLeuGlyGlySerAsnLeuLeuLeuSerAsnValThrAsp... 302
471 GCGTCCAATGCTACCTGCTCCCAATACAGTACTTCCAAAACGAAACA 520
302 ..... 302
521 GGACACAGCAAAATCCCGAGCCCTCTGGTAGTTATGCAAAATATCGCC 570
303 .AspAspSerGlyMetTyrThrCysValValTyrTyrLysAsnGluAsnI 319
571 AAGGAGCTCCCAATTCACAGGCCAGTGTACAGCCCTCATTTGAA... 617
319 leSerAlaSer.....AlaGluLeuThrValLeuValProPro 331
617 ..... 617
332 TrpPheLeuAsnHisProSerAsnLeuTyrAlaTyrGluSerMetAspI 348
618 .....TCAGTGAATGAAACACAGTT...ACCTTGAAGTAC 651
348 eGluPheGluCysThrValSerGlyLysProValProThrValAsnTrpM 365
652 TGGATAATGGA.....GCA 665
365 etLysAsnGlyAspValValIleProSerAspTyrPheGlnIleValGly 381
666 GGTGCTGAT.....GCTACTAAGGATGACGGTGTCTA 697
382 GlySerAsnLeuArgIleLeuGlyValValLysSerAspGluGlyPheTy 398
698 CTCAGGTATTTACAACTTATGACAGCAATGGTAGATACAGTGTAAAA. 746
398 rGlnCysValAlaGluAsnGluAlaGlyAsnAlaGlnThrSerAlaGlnL 415
747 .....GTGGGGGCTCTGGGAGGAGTTAACGCCACC 776
415 eulleValProLysProAlaIleProSerSerSerValLeuProSerAla 431
777 AGCGGAGAGTGATACCCAGCAGAGTGGAGCACTGTACATACCTGGCTG 826
432 ProArgAspValValProValLeuValSerSerArgPheVal..... 445
827 GATTGAGAATGATGAATAACAATGAATCCACCAAGACCTGAAATTAATA 876
446 .....ArgLeuSerTrpArgProProAlaGluAla.....L 456
877 AGGATGATGTTCACACAAGCAAGTGTGTTCAGCAGA..... 914

```

```

456 yGlyAsnIleGlnThrPheThrValPhePheSerArgGluGlyAspAsn 472
515 .....ACATCTCTCGGAGGCTCATTT.....GT 937
473 ArgGluArgAlaLeuAsnThrThrGlnProGlySerLeuGlnLeuThrVa 489
938 GGCTTCTGATGTCCTCCAAATGCT..... 959
489 lGlyAsnLeuLysProGluAlaMetTyrThrPheArgValValAlaTyrA 506
960 .....CCCAT..... 965
506 snGluTrpGlyProGlyGluSerSerGlnProIleLysValAlaThrGln 522
966 CCTGATCTCTTCCACCTGGCCAAATCACCAGCTGAAGCGGAAATTC 1015
523 ProGluLeuGlnValProGlyProValGluAsnLeuGlnAlaValSerTh 539
1016 CGGGGGCAGTCTCATTAATCTGACTTGCAGCAGCTCCTGGGGGATGATT 1065
539 rSerProThrSerIleLeuIleThrTrpGluProAla.....TyrA 554
1066 ACCATGGAACAGCTCACAGTATATCATTT.....CGAATAAGTACA 1106
554 laAsnGlyProValGlnGlyTyrArgLeuPheCysThrGluValSerThr 570
1107 AGTATTCTTGATCTCAGACAGTTCATTAATGAATCTCTTCAAGTGAATAC 1156
571 Gly..... 571
1157 TACTGCTCTCATCCCAAGGAGCAACTCTGAG...GAAGTCTTTTGT 1203
572 .....LysGluGlnAsnIleGluValAspGlyLeuSerT 583
1204 TTAACCCAGAAACATACITTTTGAATAATGCACAGATCTTTTTCATTGCT 1253
583 yrLysLeuGluGlyLeu.....LysLysPheThrGluTyrSerLeuArg 597
1254 ATTCAAGCTGTGTATAAGTGTGATCTGAATATCAGAAATATCAACATTGC 1303
599 PheLeuAlaTyrAsnArgTyrGlyProGlyValSerThrAspAspIleTh 614
1304 ACAGTATCTTTTGTATTCTCTCCACAGACTCCGCCAGACACACTAGTC 1353
614 rValValThrLeuSerAspValProSerAlaProGlnAsnValSerL 631
1354 CTGATGAACAGCTCTGCT.....CCTTGCTCT 1379
631 euGluValValAsnSerArgSerIleLysValSerTrpLeuProProPro 647
1380 AATATTATATCAACAGCACCATTCTCTGGCATTCACATT 1418
648 SerGlyThrGlnAsnGlyPheIleThrGlyTyrLysIle 660
seq_name: /SDS2/gcgdata/geneseq/geneseq/AA1991.DAT.AA13144
seq_documentation_block:
ID AA13144 standard; Protein; 1728 AA.
XX
AC AA13144;
XX
DT 04-OCT-1991 (first entry)
XX
DE Deleted in Colorectal Carcinomas.
XX
KW DCC gene; cancer; diagnosis; antibodies; tumorigenesis; neoplasm.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Protein 202..1648
FT /label= DCC

```

us-09-049-696-19.rag

371 AATGGCCTACCTCCAAAATCCAGGCAATTGCTAAGGTGGCAGCTTGGAAAT 420
+
486 sLyst 488
+
421 ACAGTCTGCAAGCAAGCTCAAAACCTTGACCTGCTAGCTGTCAGCTGGCGT 470
||||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
488 yrSerLeuLeuGlySerAsnLeuLeuLeuSerAsnValThrAsp... 503
+
471 GCGTCAATGCTACCTCGCTCCCAATTACAGTGACTTCCAAAACGAACAA 520
503 503
+
521 GGACACCAGCAAAATCCCCAGCCCTCGTAGTTTATGCAAAATATTCGCC 570
||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
504 AspAspSerGlyMetThrThrCysValValThrTyrLysAsnGluAsnI 520
+
571 AAGGACGCTCCCAAAATCTCAGGGCCAGCTGCACAGCCCTGATTTGAA... 617
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
520 leSerAlaSer.....AlaGluLeuThrValLeuValProPro 532
617 617
+
533 TrpPheLeuAsnHisProSerAsnLeuThrAlaThrGluSerMetAspII 549
+
619TCAGTGAATGGAAAAACAGTT...ACCTTGAAGACTAC 651
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
549 eGluPheGluCysThrValSerGlyLysProValProThrValAsnTrpM 566
+
652 TGGATATATGA.....GCA 665
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
566 etLysAsnGlyAspValValIleProSerAspTyrPheGlnIleValGly 582
+
663 GGTGCTGAT.....GCTACTAGGATGACGGTGCTA 697
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
583 GlySerAsnLeuArgIleLeuGlyValValLysSerAspGluGlyPheTyr 599
+
693 CTCAGGTATTTACAACTTATGACACGAAATGGTAGACAGCTGTAAAA... 746
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
593 rGlnCysValAlaGluAsnGluAlaGlyAsnAlaGlnThrSerAlaGlnL 616
+
747GTGCGGGCTCGGAGGAGTAAACGACCC 776
+
616 euIleValProLysProAlaIleProSerSerValLeuProSerAla 632
+
777 AGACGAGAGTGATACCCAGCAGAGTGAGACACTGATACCTACCTGGCTG 826
||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
633 ProArgAspValValProValLeuValSerSerArgPheVal..... 646
+
827 GATTGACAATGATGAATACATGGAATCCACCAAGACCTGAATTAATA 876
+
647ArgLeuSerTrpArgProProAlaGluAla.....L 657
+
877 AGGTGATGTTACACAAGCAAGCTGTTTTCAGCAGA..... 914
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
657 ysGlyAsnIleGlnThrPheThrValPhePheSerArgGluGlyAspAsn 673
+
915ACATCTCGGGAGGCTCATTT.....GT 937
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
674 ArgGluArgAlaLeuAsnThrThrGlnProGlySerLeuGlnLeuThrVa 690
+
938 GCTTCTCATCTCCCAATGCT..... 959
+
690 lGlyAsnLeuLysProGluAlaMetTyrThrPheArgValValAlaTyrA 707
960CCCAT..... 965
+
707 snGluTrpGlyproGlyGluSerSerGlnProIleLysValAlaThrGln 723
+
566 CQTGATCTCTTCCCACTCGGCAAAATCACCAGCTGAAGCGGAAATTC 1015
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
724 ProGluLeuGlnValProGlyProValGluAsnLeuGlnAlaValSerTh 740


```
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145931.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.

PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 23-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 23-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 114.50      Length: 299
    Ratio: 0.790        Gaps: 16
    Percent Similarity: 48.495    Percent Identity: 21.070

alignment_block:
US-09-049-696-19 x AAG43467 ..
Align_seg 1/1 to: AAG43467 from: 1 to: 386

477 AATGCTACCTCCCTCCCAATTACAGTGACTTCCAAAGCG...AACAAAGGA 523
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2 AsnAlaAlaLeuAlaThrThrAlaThrThrProValLeuArgGln 18
524 CACACAGCAAAATCCCGCCCTCTGTGTAGTTTATGCAAAATTCGCCCAAG 573
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
18 uThr.....ProLeuLeuHisTyrCysSerLeuThrThr. 29
574 GAGCCTCCCAATTC.....AGGCCAGTGTCCACAGCCCTGATT 614
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30 ..LysSerProValTyrGlnIleAsnArgValArgPheGlySerCysVal 45
615 GAATCAGTGAATGAAACACAGTTTACCTTGGAACTACTGGATAATGGAGC 664
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
46 GlnThrValSerLysLysPheLeuLysIleSerAlaSerSerGlnSerAl 62
665 AGGTGCT.....GATGCTACTAAGGATGACGGT..... 692
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
62 aSerAlaAlaValAsnValThrAlaAspAlaSerIleProLysGluMetL 79
693 .....GTCCTACTCAAGGTATTTTCAACAATTTATGACACGAATGGTAGA 734
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 ysAlaTrpValTyrSerAspTyr..... 86
735 TACAGTGTAAAGTGGGGCTCTGGAGAGGTACCGAGAGTACCGACGAGGAG 784
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 .....GlyGlyValAspValLeuLysLeuGln 95
785 A.....GTGATACCCAGCAGAGTGGAGCAGCTGTACATACCTGGCT 825
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 userAsnIleValProGlu..... 102
826 GGATTGAGATGATGAATACATGAATCCACCAAGACCTGAATTAAT 875
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 ..IleLysGluAspGlnValIleLysValValAlaAlaLeuAsn 118
876 AAGGATGATGTTCAACACAGCAAGCAAGTGTGTTTCAGCAGACATCCTCGGG 925
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119 ProValAspAlaLysArgArgGln..... 126
926 AGGCTCATTTGTGCTTCTGTATGCTCCCAATGCTCCCATACCTGATCTCT 975
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
99US-0159638.
```



```
230 GAATGGC.....ACAGTGATCGTGACAGCACCCCTGGGAAGGACACTT 273
3147 uProGlyProProLysSerLysLeuValThrAsnIleAlaLysAspSerM 3164
274 TGTTCTTTATCACCTGGACAACGACGACCTCCCAAAATCCTTCTCTGGGAT 323
3164 et...ThrValCysTrp...AsnArgProAsp.....Ser 3173
324 CCCAGTGACAGAACAGGTGGCTTCTAGTGACAAAAACACAAAT 373
3174 AspGlyGlySerGluIleGlyTyrIleValGluLysArgAspArgse 3190
374 GGCCTAC.....C 381
3190 rGlyIleArgTrpIleLysCysAsnLysArgArgIleThrAspLeuArgL 3207
382 TCCAAATCCAGGCAATGCTAAGGTGGCACTTGGAAATACAGTCTGCAA 431
3207 euArgValThrGlyLeuThrGluAspHisGluTyrGluPheArgValSer 3223
432 GCAAGCTCA.....CAACCTTGACCTGACTGTCAAGTC 466
3224 AlaGluAsnAlaAlaGlyValGlyGluProSerProAlaThrValTyrTy 3240
467 CCGTGGCTCCCAATGCTACCTCGCTCCCAATACAGTGACTTCCAAACGA 516
3240 rLysAlaCysAspProAlaPheLysProGlyProProThrAsnAlaHisI 3257
517 ACAAGGACACAGCAAAATCCCGACCTCTGCTGATTTATGCAAAATAT 566
3257 leValAspThrThrLysAsnSerIleThrLeuAlaTrpGlyLysProIle 3273
567 CGCAAGAGAGCTCCCAATTCCTAGGCCCAAGTGTACAGCCCTGATTGA 616
3274 TyrAspGlyGlySerGluIleLeu..... 3281
617 ATCAGTGAATGAAAAACAGTACTCTGGAATCTACTGGATAATGAGCAG 666
3282 .....GlyTyrValValGluIleCysLysAlaAspGluGluGluT 3295
667 GTGCTGATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTTC..... 710
3295 rpGlnIleValThrProGlnThrGlyLeuArgValThrArgPheGluIle 3311
711 ACACTTATGACAGCAATGCTAGATACAGTGTAAAGTGGCGGTCTGGG 760
3312 SerLysLeuThrGluHisGlnGluTyrLysIleArgValCysAlaLeuAs 3328
761 AGGAGTTAACGCAGCAGCAGAGAGTGTATACCCAGCAGAGTGAGCAC 810
3328 nLysValGlyLeuGlyGlu.....AlaTr 3336
811 TGTACATACCTGGCTGGATTGAG...AATGATGAATACAAATGGAATCCA 857
3336 hrSerValProGlyThrValLysProGluAspLysLeuGlu..... 3349
858 CCAAGACCTGAATTAATAGATGATGTTCAACACAGCAAGTGTGTTT 907
3350 ...AlaProGluLeuAspLysSerGluLeuArgLysGlyIleVal... 3364
908 CAGCAGAACATCTCGGGAGGCTATTGTGGCTTCTGATGTCCTCA... 953
3365 .....ValArgAlaGlyLysSer...AlaArgIleHisIlePropheL 3378
954 ...AATGCTCCCATACCTGATCTTTC.....CCACCTGGCCAAATC 992
3378 ysGlyArgProThrProGluIleThrTrpSerArgGluGluGlyGluPhe 3394
993 ACCGACCTGAAGGGGAAATTCAGCGGGCAGTCTCTATTATCTGCTTG 1042
3395 ThrAsp...LysValGlnIleGluLysGly.....ValAsnTyrThrGl 3408
1043 GACAGCTCTGGGGATGATTATGACCATGGAGACAGCTCACAAAGTATATCA 1092
```

```
3408 nLeuSerIle...AspAsnCysAspArgAsnAspAlaGlyLysTyrIleL 3424
1693 TTCGAATAAGTACAAAGTATTCTTGATCTTCAGAGACAAGTTCATCAATCT 1142
3424 euLysLeuGluAsnSer...SerGlySerLysSerAlaPheValThrVal 3439
1143 CTTCAGTGAATACTACTGCT.....CTCATCCCCAAAGGAAGC 1180
3440 LysValLeuAspThrProGlyProGlnAsnLeuAlaValLysGluVa 3456
1181 CAACTCTGAGGAAGTCTTTTG..... 1202
3456 lArgLysAspSerAlaPheLeuValTrpGluProProIleIleAspGlyG 3473
1202 ..... 1202
3473 lYAlaLysValLysAsnTyrValIleAspLysArgGluSerThrArgLys 3489
1203 .....TTTAAACCCAGA 1213
3490 AlaTyrAlaAsnValSerSerLysCysSerLysThrSerPheLysValGl 3506
1214 AACATTAATCTTTGAAAATGCGACAGATCTTTTCATTCCTATTCAGGCTG 1263
3506 uAsnLeuThr.....GluGlyAlaIleTyrTyrPheArgValMetAlaG 3521
1264 TTGATAAGTCTGATCTGAAATCAGAAATATCCAACTATTCACGAGTATCT 1313
3521 LuAsnGluPheGlyValGlyValProValGluThrValAspAlaValLys 3537
1314 TTGTTTATCTCTCCACAGACTCCGCCAGACACACCTAGTCTCTGAT..... 1358
3538 AlaAlaGluProProSerProGlyLysValThrLeuThrAspValSe 3554
1359 GAAAGCTCTGCTCTCTGCTCTAATATTCATATCAACAGCACCATTCTCTG 1407
3554 rGlnThrSerAla..... 3558
1408 GCATTCACATTTTAAAAATATTGTGAAG 1436
3559 .....SerLeuMetTrpGlu 3563
seq_name: /SDS2/gc9data/geneseq/geneseq/AA1992.DAT:RAR26983
seq_documentation_block:
ID RAR26983 standard; Protein; 637 AA.
XX
AC RAR26983;
XX
DT 11 FEB-1993 (first entry)
XX
DE (FRP51)-ETA fusion protein.
XX
KW Monoclonal antibody; light chain; heavy chain; tumour; c-erbB-2;
KW variable region; ETA.
XX
OS Pseudomonas aeruginosa.
XX
FH Key Location/Qualifiers
FT Peptide 1..21 /label= ompA_signal_peptide
FT Peptide 22..29 /label= FLAG_peptide_and_enterokinase_cleavage_site
FT Domain 33..152 /label= FWP51_heavy_chain_variable_domain
FT Peptide 153..167 /label= Linker
FT Domain 168..274 /label= FRP5_light_chain_variable_domain
FT Protein 276..397 /label= ETA_252-613
XX
```


CC Robo (roundabout) on its surface. The method comprises modulating the
CC effective amount of Comm polypeptide in contact with the cell, where the
CC amount of expressed active Robo is specifically modulated inversely with
CC the modulation of the effective amount of Comm in contact with the cell.
CC The method is used to modulate the amount of active Robo expressed on a
CC cell. The method can be used to screen for agents that modulate Robo:Comm
CC interactions. This is particularly useful for modulating nerve cell
CC function.

SQ Sequence 1395 AA;

alignment_scores:

Quality: 111.00 Length: 472
Ratio: 0.483 Gaps: 21
Percent Similarity: 48.729 Percent Identity: 21.186

alignment_block:

US-09-049-696-19 x AAY13563 ..

Align seg 1/1 to: AAY13563 from: 1 to: 1395

```
144 GCCCTTTTCATCAGGAATGGAGCTGCTCTCAGCGCTCCATCCAGCTTGA 193
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
484 AlaValGlnAlaGlyAsn.ArgTyrSerIleIleGlnGlySerLeuA 500
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 GAGTAAGGGATTAAACC.....TCCAGAACAGCC 222
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
500 rgValAspAspLeuGlnLeuSerAspSerGlyThrTyrThrCysThrAla 516
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
223 AGTGATGAATGGCAGTGTGCGGACAGCAGCCGTGGGAAGAGCACT 272
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
517 SerGlyGluArgGlyGluThrSerIleIleAla.....ThrLe 529
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
273 TTGTTCTTATCAGTCGACACGAGCGCTC.....CCCA 307
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
529 uThrValGluLysProGlySerThrSerLeuHisArgAlaAlaaspPro 546
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
308 AATCTCTCTCGGGATCCCA.....GTGACACAGA 336
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
546 erThrTyrProAlaProGlyThrProLysValLeuAsnValSerArg 562
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
337 AGCAGGTGGCTTGTACTGACAAAAACACCAAAATGGCTACTCCAA 386
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
563 ThrSerIleSerLeuArgTyrAlaLysSerGlnGluLysProGlyAla 579
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
387 ATCCAGCATGTCTAAGTGGCACTTGGAAATACAGTCTGCAAGCAAG 436
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
579 lGly...ProIleIle.GlyTyrThrValGluTyrPheSerProAspLeu 594
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
437 CTCACA.....AACCTTGACCTGACTGTACAGTCCC 468
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
595 GlnThrGlyTrpIleValAlaAlaHisArgValGlyAspThrGlnValTh 611
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
469 GTGCGTCCAATGTACTACCTGCTCCCAATACAGTACTCCAAAACGAAC 518
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
611 rIleSerGlyLeuThrProGlyThrSerTyrValPheLeuValArgAlaG 628
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
519 AAGGACACCAACAAATCCCGAGCCCTCTGGTAGTT..... 554
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
628 luAsnThrGlnGlyIleSerValProSerGly.LeuSerAsnValIleLy 644
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
555 .....TATGCAAAATATCCCAAGGAG 576
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
644 sThrIleGluAlaAspPheAspAlaAlaSerAlaAsnAspLeuSerAla 661
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
577 CCTCCCAATTTCTAGGCCAGTGTACAGCCCTGATTGAA.....TCA 620
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
661 laArgThrLeuLeuThrGlyLysSerValGluLeuIleAspAlaSerAla 677
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
621 GTGAATGAAAAACAGTTACCTTGGATATCTGGATAATGGAGCAGGTGC 670
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
678 IleAsnAlaSerAlaValArgLeuGluTrpMetLeuHis....ValSerAl 693
```

```
671 TGATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCACAACCTTATG 720
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
693 aAspGluLysTyrValGluGlyLeuArgIleHisTyrLysAspAlaSerV 710
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
721 ACACGAATGGTAGATAC...AGTGTAAGAAGTGGGGCTCTGGAGAGGAGT 767
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
710 alProSerAlaGlnTyrHisSerIleThrValMetAspAlaSerAlaGlu 726
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
768 AACGACGACGAGGAGATGATATACCCGACGAGGTGGAGCACTGTACAT 817
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
727 SerPheValValGlyAsnLeuLysLysTyrThrLysTyrGluPhePheLe 743
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
818 ACCTGGCTGGATTGAGATGATGAATAACATGGAATCCACCAAGACCTG 867
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
743 uThrProPheGluThrIleGluGlyGln.....ProS 755
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
868 AATTAATAAGGATGATGTTCAACACAAGCAAGTGTGTTTCAGCAGAACA 917
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
755 erAsnSerLysThrAlaLeuThrTyrGlu..... 764
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
918 TCCTCGGAGGCTCATTTGTGGCTTCTGATGTCCTCAAAATGCTCCCATACC 967
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
765 .....AspValProSerAlaPro...Pr 771
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
968 TGATCTCTTCCACCTGGCCAAATCACCGACCTGAAGCGGGAATTCAGG 1017
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
771 oAspAsnIleGlnIleGlyMetTyrAsnGlnThrAlaGlyTrpValArg. 787
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1018 GGGCGAGCTCATTAATCTGACTTGGCAGCTCCTGGGGATGATTATGAC 1067
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
788 .....TrpThrProProSerGlnHisHis 796
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1069 CATGGAACAGCTCACAGTATATATCATCAATAAGTACAAAGTATTCTTGA 1117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
797 AsnGlyAsnLeuTyrGlyTyrLysIleGluValSerAlaGly...AsnTh 812
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1118 TCTCAGACAGAAGTTCATGAATGATCTCTCAAGTGAATACTACTGCTCTCA 1167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
812 rMetLysValLeuAlaAsnMetThrLeuAsnAlaThrThrThrServall 829
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1168 TCCCAAGGAAGCCAACTCTGAGGAAGTCTTTTGTTTTAAACCAAGAAAC 1217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
829 euLeuAsnAsnLeuThrThrGlyAlaValTyrServallArgLeuAsnSer 845
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1218 ATTACTTTTGAATAATGGCACAGATCTTTTCATTTGCTATTTCAGGCTTGA 1267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
846 PheThr.....LysAlaGlyAs 851
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1268 TAGGTCTGATCTGAATCAGAAATATCCACATTCGACGAGTATCTTTGT 1317
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
851 pGlyProTyrSerLysPro.....IleSerLeup 861
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1318 TTATT.....CCTCCACAGACTCCGCGCAGAGACA 1346
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
861 heMetAspProThrHisHisValHisProProArgAlaHisProSerGly 877
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1347 CCTAGTCTGATGAACCTCTGCTCTCTTGTCTCTTAATATTCATATCAACAG 1396
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
878 ThrHisAspGlyArgHisGluGlyGlnAspLeuThrTyrHisAsnAsnGl 894
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1397 CACATTCTCT 1406
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
894 yAsnIlePro 897
```

This Page Blank (uspto)

OM of: US-09-049-696-18 to: SPTREMBL_17:* out_format : pfs

Date: Mar 30, 2002 2:47 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09049696/runat_28032002.145238.2053/app_query.fasta.1.12579
-DB=SPTREMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGNL_1_705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-18

Query length: 2813

Database: SPTREMBL_17.*

Database sequences: 473505

Database length: 146272329

Search time (sec): 805.760000

score_list:

Sequence	Strd Orig	ZScore	Escore	len	Documentation
sp_mammal:Q29282	+ 278.00	369.87	2.3e-12	109	! Q29282 sus scrofa (pig). unknown
sp_plant:Q92V10	+ 153.00	187.43	0.0053	689	! Q92V10 arabidopsis thaliana (ma
sp_bacteria:Q9PLZ9	+ 150.50	181.35	0.0085	928	! Q9PLZ9 campylobacter jejuni. pc
sp_bacteria:Q9L950	+ 149.50	172.24	0.0119	2147	! Q9L950 pseudomonas putida. out
sp_bacteria:Q9AES2	+ 144.50	170.77	0.0249	1234	! Q9AES2 clostridium ramosum. 19
sp_bacteria:Q9RNI2	+ 144.50	168.61	0.0260	1357	! Q9RNI2 haemophilus influenzae. 19
sp_bacteria:Q9ZD91	+ 143.50	163.51	0.0333	2340	! Q9ZD91 rickettsia prowazekii.
sp_bacteria:P71401	+ 141.50	160.81	0.0469	2353	! P71401 haemophilus influenzae.
sp_invertebrate:Q9V7P4	+ 140.50	173.99	0.0412	424	! Q9V7P4 eufolliculina uhligi.
sp_invertebrate:Q9V3J6	+ 140.50	170.47	0.0443	721	! Q9V3J6 drosophila melanogaste
sp_bacteria:P77847	+ 140.50	168.30	0.0463	911	! P77847 caldocellum saccharolyti
sp_bacteria:Q9PHN8	+ 140.00	166.98	0.0511	978	! Q9PHN8 campylobacter jejuni. pc
sp_bacteria:Q52781	+ 139.50	165.13	0.0570	1112	! Q52781 campylobacter fetus. su
sp_invertebrate:Q9N5K0	+ 139.50	164.40	0.0579	1203	! Q9N5K0 caenorhabditis elegans
sp_vertebrate:Q9W6V5	+ 139.00	162.29	0.0649	1406	! Q9W6V5 gallus gallus (chicken)
sp_bacteria:Q48031	+ 138.50	160.80	0.0718	1536	! Q48031 haemophilus influenzae
sp_invertebrate:Q9U3G8	+ 138.00	167.67	0.0869	683	! Q9U3G8 caenorhabditis elegans
sp_archaea:Q26996	+ 136.00	164.82	0.0944	698	! Q26996 methanobacterium thermo
sp_bacteria:Q99U54	+ 135.50	143.13	0.1585	6713	! Q99U54 staphylococcus aureus s
sp_rodent:Q55005	+ 135.00	155.50	0.1319	1651	! Q55005 rattus norvegicus (rat)
sp_bacteria:P76017	+ 134.50	159.93	0.1293	955	! P76017 escherichia coli. from h
sp_bacteria:Q9X6M3	+ 134.50	155.10	0.1428	1605	! Q9X6M3 salmonella typhi. proli
sp_vertebrate:Q90793	+ 134.00	157.36	0.1463	1173	! Q90793 gallus gallus (chicken)
sp_human:Q10466	+ 134.00	128.25	0.2666	26926	! Q10466 homo sapiens (human).
sp_bacteria:Q9XCJ4	+ 133.50	151.58	0.1770	2035	! Q9XCJ4 salmonella typhimurium.
sp_invertebrate:Q18005	+ 133.00	163.86	0.1476	505	! Q18005 caenorhabditis elegans

seq_name: sp_mammal:Q29282

seq_documentation_block:

ID	Q29282	PRELIMINARY;	PRT;	109 AA.
AC	Q29282;			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-WAY-1997	(TREMBLrel. 03, Last annotation update)		
DE	UNKNOWN PROTEIN (FRAGMENT).			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.			
OX	NCBI_TaxID=9823;			

RN	[1]	SEQUENCE FROM N.A.
RP		TISSUE=SMALL INTESTINE;
RX		MEDLINE=96327607; PubMed=8672129;
RA		Witwer A.K., Fredholm M., Davies W.;
RT		"Evaluation and characterization of a porcine small intestine cDNA
RL		library: analysis of 839 clones.";
DR	EMBL; F15082; CAA23338.1; -	Mamm. Genome 7:509-517(1996).
FT	NON_TER	1
FT	NON_TER	109
SQ	SEQUENCE	109 AA; 12048 MW; 67FCD61718B151CD CRC64;

alignment_scores:

Quality:	278.00	Length:	99
Ratio:	3.390	Gaps:	4
Percent Similarity:	82.828	Percent Identity:	67.677

alignment_block:

US-09-049-696-18 x Q29282 ..

Align seg 1/1 to: Q29282 from: 1 to: 109

937 AGAATTGTTGTTAGTCTTCGACAAATCTGGAAGCATGGCGACTGGTAA 986

1 ArgileValCysLeuValLeuAspLysSerGlySerMetThrValGlyGI 17

987 CCGCTCAATCGACTGAATCAAGCAGCGCAGCTTTCTCTGTCGACAGAC 1036

17 YArgLeuLysArgLeuAsnGlnAlaGlyLysLeuPheLeuLeuGlnThrV 34

1037 TTGAGCTGGGTCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1086

34 alGluGlnGlyAlaTirpValGlyMetValAlaPheAspSerAlaAlaTyr 50

1087 GTCAAAAGTGAACATACAGATAAAGCAGTGGCAGTGAC...AGGCAC 1133

51 ValLysSerGluLeuValGlnAlaGlnAlaAspSerAlaAlaAsp***ArgGlnCy 67

1134 ACTGCCAAAAGATGACTCTGCAGCAGCTTCAAGAGGAGGAGGAGGAGGAG 1180

67 sLeu.ProAlaAlaTyrProGlnGlnPro.Arg***GluProSerIleLe 83

1181 GCACGGGCTTGCATGGCGACTTACTGTGATTAGGAGAAATA 1223

83 uLeuGlyAlaSerLeuArgIleTyrCysAspGlnGluGluIle 97

seq_name: sp_plant:Q92V10

seq_documentation_block:

ID	Q92V10	PRELIMINARY;	PRT;	689 AA.
AC	Q92V10;			
DT	01-WAY-1999	(TREMBLrel. 10, Created)		
DT	01-WAY-1999	(TREMBLrel. 10, Last sequence update)		
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)		
DE	RETROTRANSPON- LIKE PROTEIN.			
GN	T7F6.14.			
OS	Arabidopsis thaliana (Mouse-ear cross).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.			
OX	NCBI_TaxID=3702;			
RN	[1]	SEQUENCE FROM N.A.		
RP		STRAIN=CV, COLUMBIA;		
RC	Kounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,			
RA	Shen M., Renning C.M., Fraser C.M., Somerville C.R., Venter J.C.;			
RT	"Arabidopsis thaliana chromosome II BAC T7F6 genomic sequence.";			
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.			
CC	- SIMILARITY: CONTAINS A RING-TYPE ZINC FINGER.			
DR	EMBL; AC005770; AAC79610.1; -			
DR	InterPro; IPR02035; VWFA.			

517 ...GTATTGACGAGTCAATAATGATGAGAAATTCATTATCAATGG 563
124 SerIleLeuGluTyrGlyGlyPheIleGluHisIleIleAsnThrG1 140
564 AAGATAACAA...GCAGTAAGATGTTACAGCAGGTATTACTG 601
140 yThrIleSerAsnAsnValGlyIleGlySerAlaTyrGlyValPheG 157
602 GTACAATGTAGTAAGAAGTGTGAGGAGCAGCTGTATACACAAA... 648
157 lyThrLeuThrIleLys...AspGlyGlyMetValTyrGlyLysTyr 171
648 648
172 SerAlaIleGlyValGlyArgSerGlnThrLeuGlyAspLeuTyrIleAs 188
649 ...AGATGCACATTCATAAAGTAAGTAAGCAGGACTCTATGAAAAGGATGTG 694
188 pGlyArgSerAsnAsnGlyThrValSerGlyIleTyrSerCluGluHisG 205
695 AGTTTGTCTCCATCCCGCCAGCAGGAGAGAGCTTCTATATGTTTGC 744
205 lyIleLeuLeuGluAsn..... 210
745 CAACATGTTGATTCTATAGTTGAATTCGTACAGAACAAACCAACAA 794
210 210
795 AGAAGCTCCAAACAGCAAAATCAAAATGCAATCTCCGAGCAGCATGGG 844
211AsnSerArgThrGlnLysIleGluLeuLysAsn...GlyG 223
845 AAGTGATCCGTGATCTGAGGACTTTAAGAAACCACTCTCTATGACAACA 894
223 lyIleIleLysGlyAsnIleAsp..... 230
895 CAGCCACCAATCCACCTTCTCATGTCGAGATTGGACAAAGAATTGT 944
231GlyIleArgLeuI1 235
945 GTGTTTAGTCCTTGACAAATCTGGAAGCA 973
235 eAsnSerAlaSerLeuSerGlyGluMetIleLeuSerGlyGluGlySerA 252
974 TGGCGACTGTAAACGC...CTCAATCGACTGAATCAAGCAGCCAGCTT 1020
252 rgValGluGlyGlyArgGlyValGlyIleLeuAsnArgSerGlyLysIle 268
1021 TTCCTGCTGCAGACAGTTGAGCTGGGTCTCTGGCTT 1056
269 GluGlySerIleLysValGluAspGlyAlaThrValThrAlaThrSerAs 285
1057GGGATGGTGACA..... 1068
285 nArgAlaIleAlaAsnSerGlySerGlySerIleThrGlyGlyIleThrV 302
1068 1068
302 alSerGlyLysAsnThrLysLeuGluGlyAsnIleIleAsnThrGlyAsn 318
1069TTTGACAGTGTGCCCATGTACAAAG 1094
319 AlaSerIleGlySerAspIleLysIleGluGlyAlaLysValGluG1 335
1095 TGAATCTC.....ATACAGATAA 1111
335 yGlyLeuValAsnGlnGlyAsnGlySerIleSerGlySerValGlnVals 352
1112 ACAGTGGCAGTGACAGGACACACTGCCAAMAGATTACCTGCAGCAGCT 1161
352 erGlyGlySerSerIleAspSerIleThrAsnGluGlyAsnGlyAlaIle 368
1162 TCAGGAGGAGCTCCATCTGC.....AGCGGGCTTCGATCGGCATTTC 1205

369 SerGlySerIleThrValTyrLysAspSerLysLeuAspSerIleThrAs 385
1206 TGTGATTAGGAAGAAATATCCAACTGATGATCTGAAATTCGTCTGCTGA 1255
385 nThrSerThrSerSerThrGlyIleSerGlySer.....IleT 398
1256 CGGATGGGGAAGACAACACTATAAGTGGGTCTTTTAACGAGGTCAACAA 1305
398 hrAsnAsnSerAspAsnLeu.....GluIleSerAsn 409
1306 AGTGGTCCATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGA 1355
410 SerGlyAsnIle..... 413
1356 ACTAGAGGAGCTGTCCAAAATGACAGGAGGTTTA.....CAGACATATG 1399
414GlyGlyLysIleGluSerThrGlyS 422
1400 CTTGAGATCAAGTT.....CAGAACAATGGCCTCATTTGATGCTTTT 1440
422 erAlaAspMetValIleSerAsnSerAsnGlyGlyThrIleSerGly... 437
1441 GGGCCCTTTTCATCAGGAATGGAGCTCTCTCAGCGCTCCATCCAGCT 1490
438 GlyIleSerSerGlySerGlySerThrSerIleSerAsn..... 451
1491 TGAGATGAAGGATTAAACCTCCAGAACAGCAGTGGATGAATGGCAGC 1540
452SerGlnGlySerThrIleAsnAsnGlyIleThrValSerGlySerA 467
1541 TGATCTGTGGAC.....AGCACCGTGGGAAAGGACACACTTTGTTT 1578
467 laGlnValGluIleSerAsnGlnGlySerValGlyLysAsp..... 480
1579 CTTATACCTGGACACGCCACCTCCCAATCTCTCTGGGATCCAG 1628
480 480
1629 TGSACAGAAGCAAGTGGCTTTTGTAGTGACAAACCAACCAAAATGGCCT 1678
481GluAsnGlyAsnThrValThrAsnAsn..... 489
1679 ACCTCCAAATCCAGGCAATTCCTAAGTTGGC.....ACTTGGAAATAC 1722
490GlySerGlySerValGlyIleLysAspTrpLeuVal 501
1723 AGTCTGCAAGCAAGCTCACAAACCTTGACCTGACTGTGACG.....TC 1766
502 SerThrAspLysAsnThrGlyLysLeuAsnThrValIleGlyGlySe 518
1767 CCGTGCCTCCAATGCTACCTCCCTGCCAATTTACAGTGACTTCCAAAACGA 1816
518 rArgAlaPheAsnValLysValGluAsnIleThrVal...AspGlnSerA 534
1817 ACAAGGACACAGCAAAATTCGCCAGCCCTCTGTTAGTTTATGCAAAATATT 1866
534 snValAspLeuGluGluLeu..... 540
1867 CGCCAAGGAGCTCCCAATTCAGGCGCCAGTGTACAGCCCTGATTGA 1916
541AsnAspIleAsnAsnIleIleSe 548
1917 ATCAGTGAATGGAAAAACAGTTACTTTGGAATCTGATATGAGCAG 1966
548 rGlyValAsnGlnAsnAsnIleGly...AsnIleGlyThrAsnGlySerG 564
1967 GTGCTGTGCTACTAAGGATGACCGTGTCTACTCAAGGTATTTCACAACT 2016
564 lyGluIleSerLeuSerPheAspProIleThrGlyLysLeuThrThrAsp 580
2017 TATCACACGAATGTAGATACAGT...GTAAAAGTGGCGCTCTGGGAGG 2063

```
581 PheAsnLeuAsnAlaSerIleSerGlyAlaThrPheArgSerLeuIleSe 597
2064 AGTTACGCGAGCAGCGAGAGTGTATACCCAGAGTGGAGCACTGT 2113
597 rThrThrSerArgArgSerThrPheIle
2114 ACATACCTGGCTGGATGTGAGATGATGAATACAAATGGAATCCACCAAGA 2163
607 .....AspAsnValMetGlyAsn..... 612
2164 CTTGAATTAATAAGGATGATGTCACACAAGCAAGTGTGTTTCAGCAG 2213
613 .....SerMetGlnSerPheAlaLeuAlaSe 621
2214 AACATCTCGGGAGGCTCTTTGTGCTCTCTGATGTCCCAAT.....G 2257
621 rSerSerLysSerGlnSerIleAlaMetSerGluLysGlyAsnLeuTyra 638
2258 CTCCCATACCTGATCTCTTCCCACTGGCCAAATCACCCAGCTGAAGCG 2307
638 laAspAlaSerAspTyr.....IleLysSer 646
2308 GAAATTCAGGGGAGTCTCATTAATCTGACTTGGACAGCTCTCTGGGA 2357
647 AspLeuAsnGlySer..... 652
2358 TGATTATGACCATGGAACAGCTCACAGATATATCATTCGAATAAGTACAA 2407
653 ....TyrGlySerAsnLysGluHisSerLeuPheIleLeuProTyrThrS 668
2408 GTATCTTTCATCTCAGACAGCAAGTCAATGAATCTCTTCAAGTGAATCT 2457
668 erSerGlnAsnValGluLeuSerLeuAsnGluSerLysGlyHisThr 684
2458 ACTGCTCTCATC.....CCAAAGGAGCCCACTCTCAGGA 2492
685 LysGlyThrIleIleGlyTyrSerThrLeuLysAspSerGlyIleTyrGI 701
2493 AGTCTTT.....TTGTTTAAAC 2509
701 yValTyrAlaGlyTyrGluAspThrLysMetGlySerThrTyrPheAspI 718
2510 CAGAAACATTTCTTT..... 2526
718 leAsnAsnArgThrTyrTyrAlaGlyLeuLysTyrPheAsnThrLeuPhe 734
2527 .....GAAATGGCACAGATCTTTTCATCTGCTATTACG.....GC 2561
735 ThrThrGluLysGlyGlnGluValTyrIleLysAlaGlnGlyLysAlaAl 751
2562 TGTGTATAGTGCATCTGAATCAGAAATCAGAAATATCCAC 2598
751 aLeuIleLysAsnAspLeuThrGluLysIleGlyAsn 763
```

seq_name: sp_bacteria:Q9L950

```
seq_documentation_block:
ID Q9L950 PRELIMINARY; PRT: 2147 AA.
AC Q9L950;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE OUTER MEMBRANE-LIKE PROTEIN.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20225830; PubMed=10762233;
RA Espinosa-Urgel M., Salido A., Ramos J.L.;
RT "Genetic analysis of functions involved in adhesion of Pseudomonas putida to seeds."
```

```
RL J. Bacteriol. 182:2363-2369(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Espinosa-Urgel M., Salido A., Ramos J.L.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EM3L; AF182515; AAF69025.1; -
DR InterPro; IPR001343; Hemlyso_Ca_bind.
DR PRINTS; PR00313; CABNDNGRPT.
SQ SEQUENCE 2147 AA; 211585 MW; BD7FC3B460FE2D6B CRC64;

alignment_scores:
    Quality: 149.50      Length: 774
    Ratio: 0.430        Gaps: 37
    Percent Similarity: 44.961      Percent Identity: 19.897

alignment_block:
US-09-049-696-18 x Q9L950 ..
Align seg 1/1 to: Q9L950 from: 1 to: 2147

844. GAAGTGATCCGCTGATTCGAGGACTTTAAGAAACCACCTCTATGACAAC 893
337 GlnValSerAlaThrGlyProAspGlyThrValSerLeuProAlaThrLe 353
894 ACAGCCCAAAATCCACCTCTCTCATTCTCCAGATTGGACAAGAATTG 943
353 uGlnAlaProAspThrThr.....AlaProGlnProLeu 365
944 TGTGTTTATGCTCTCACAATACTGGAAGCATCGCGACTGGTAAC..... 987
365 hrAsnLeuValLeuSerAsnAspGlyLeuValLeuSerGlyArgGlyGlu 381
988 .....CGCCTCAATCGACTGAATCAAGCAGCGCAGCTTTCTCTGCT 1028
362 AlaGlyAlaThrValThrValLeuGlyGluGlyGly..... 393
1029 GCAGACAGTTGAGCTGGGTCTCTGGTGGGATGGTGGATTTGACAGTG 1078
394 .....IleGluLeuGlyThr.....AlaValValAspAlaAsnGlyA 406
1079 CTGCCCATGTACAAAGTGAACCTCATACAGATAAAC..... 1113
406 spPheAsnValAspLeuSerValAlaGlnIleAsnGlyGluGlnLeuSer 422
1114 .....AGTGGCAGTGACAGGACACACTCGCCCAAAAG 1145
423 AlaSerGlnSerAspValAlaGlyAsnGluSerAsnSerValSerLeuTh 439
1145 ATTACT.....GCAGCAGCTTCAG 1165
439 rAlaProAspArgValAlaProAspAlaAlaGlyAsnLeuAlaPheAlaG 456
1165 GAGGACGCTCCATCTGACGCGGCTTCGATCGGCATTACTGTGATTAGG 1215
453. lYgLyGlySerLeuLeuAsnGlyThrGlyGluAlaGlyAlaAlaValArg 472
1216 AAGAATATCCCACTGATGATCTGAATTTGTGCTGACGCGATGGGA 1265
473 .....ValLeuAlaAlaAsp..... 477
1265 AGACAACACTATAAGTGGGTGCTTTAACGAGGTCAAACAAGTGGTCCA 1315
478 .....GlySerV 480
1315 TCATCCACACAGTCGCTTTGGGCGCTCTGACGCTCAACAAGTAGAGGAG 1365
480 aLeuGlySerAlaThrValArgAlaAspGlyThrPheGlnValAsnLeu 496
1365 CTGCTCCAAAATGACAGGAGT.....TTACAGACATATGCTTCAGATCA 1409
497 ValSerProGlnAlaAsnGlyGlnGlnValGlnValValLeuAlaAspAl 513
```


1410 AGTTCAGACAAATGGCTCATGTGCTTTTGGGGCC..... 1446
::: |||::: |||::: |||::: |||::: |||::: |||
513 aAlaGlyAsnGlnSerAlaAlaSerAlaIleThrAlaProAspThrThr 530
1447CTTTCATCAGGAATGGAGCTGCTCTCAGCGCTCCATCAGCTT 1491
::: |||::: |||::: |||::: |||::: |||::: |||
530 roProSerAlaProGlyGluLeuAlaValSerGlyAspGlyValThrLeu 546
1492 GAGAGTAAGGA.....TTAACCTCCAGAACACCCAGTG 1526
::: |||::: |||::: |||::: |||::: |||::: |||
547 ThrGlyArgGlyGluValGlyAlaArgValThrValThrAsnProAla... 562
1527 GATCAATGGCACATGCTGGACAGCAGCTGGGAAGGACACTTGT 1576
|||::: |||::: |||::: |||::: |||::: |||::: |||
563GlyGluGluIleGlySerAlaThrValAlaSerGlyHisP 577
1577 TTCTTATCACCTGG..... 1590
|||::: |||::: |||::: |||::: |||::: |||::: |||
577 heThrIleThrLeuSerProAlaLeuGlyAsnAlaGluLeuLeuSerPhe 593
1591ACAACGCCAGCC 1601
594 ThrGlnAlaAspAlaAlaGlyAsnValSerAlaValAlaThrLeuGlnTh 610
1602 TCCCAATCTCTCTGGGATCCAGTGGACAG..... 1635
|||::: |||::: |||::: |||::: |||::: |||::: |||
610 rProAspPheThrProProAspProLeuThrGlnValValIleAsnAla 627
1636AAGCAAGGTGGCTTGTAGTGGAC 1659
::: |||::: |||::: |||::: |||::: |||::: |||
627 spGlySerValValSerGlyLeuGlyGluAlaGlyAlaThrValIleVal 643
1660 AAAACACCAAAATGGCTACTCTCAATCCAGGCATTTCTAAGGTGG 1709
::: |||::: |||::: |||::: |||::: |||::: |||
644 SerAsnAlaAlaGlyThrValLeuGlyThrGlyIleValLeuAlaAspG 660
1710 CACTTGGAAATACAGTCTG.....CAGCAAGCTCACAACTTGA 1750
|||::: |||::: |||::: |||::: |||::: |||::: |||
660 ySerPheSerValGluLeuSerProAlaGlnIleAsnAsnGlnThrLeu 677
1751 CCCTGACTGTACGTCGCGTCCCAATGCTACCTG..... 1788
::: |||::: |||::: |||::: |||::: |||::: |||
677 erValLeuGlnAlaAspProProGlyAsnValSerLeuProValAsnVal 693
1789CCTCCAAATFACAGTACTTCC...AAACGAA 1817
|||::: |||::: |||::: |||::: |||::: |||::: |||
694 GlnAlaProAspLeuThrProProAlaAlaAlaThrGluLeuArgLeu 710
1818 CAAG.....GACACAGCAAAATCCCGACCTCTGG 1849
|
710 nAlaAlaGlyAspGluLeuGlyGlyValGlyGluValGlyAlaThrVal 727
1850 TAGTTTATGCAAAATATTCGCCAAGAGCCTCCCAATTTCTCAGGCCAGT 1899
|||::: |||::: |||::: |||::: |||::: |||::: |||
727 rgValTyrValGlyThrGlnIleGlyThrGlyThrValGlyAlaAsn 743
1900 GTCACAGCCCTGATTGAA.....TCAGTGAATGGAACACAGT 1937
|||::: |||::: |||::: |||::: |||::: |||::: |||
744 GlyThrPheIleValGluLeuAspAlaAlaGlnLeuAspGlyGlnVal 760
1938 TACCTTGAAGTACTGGATATGAGCAGGTGCTGATGCTACTAAGGATG 1987
::: |||::: |||::: |||::: |||::: |||::: |||
760 uGlnValThrLeu.....ThrAspAlaThrGlyAsnI 771
1988 ACGGTGCTACTCAAGGTATTTCAACTATGACAGCAATGGTAGATAC 2037
::: |||::: |||::: |||::: |||::: |||::: |||
771 leSerProIleSerSer...ValThrAlaValAspThrThrProProAla 786
2038 AGTCTA.....AAAGTGGCGGCTCTGGAGGAGT 2066
::: |||::: |||::: |||::: |||::: |||::: |||
787 ThrValValAlaThrIleAsnAspSerGlyThrGlnValIleGlyThrAl 803
2067 TAACGCCACCGAGAGTGTATACCCACAGCAGTGGAGCAGCTGTACA 2116

803 aGluAlaGlyThrArgValThrValValAsnAsnGlyAspLeuLeu 820
2117 TACCTGGCTGATTGAGAATGAT.....GAAATACAATGGAATCCA 2157
::: |||::: |||::: |||::: |||::: |||::: |||
820 lyGlnValThrValGluProAspGlyThrPheValLeuAspLeuAsn... 835
2158 CCAAGACCTGAAATTAATAGGATGATTTCACACACAGCAAGTGTGTTT 2207
|||::: |||::: |||::: |||::: |||::: |||::: |||
835 ...LeuProGlnIleAsnGlyGluThrLeu..... 844
2208 CAGCAGACATCTCCGGAGGCTATTGTGGCTCTCGATGTC..... 2250
|||::: |||::: |||::: |||::: |||::: |||::: |||
845ThrIleAlaGlnAspAlaThrGly 854
2251 ..CCAATGCTCCCAT.....CCTGATCTCTCCACCTGGC 2286
|||::: |||::: |||::: |||::: |||::: |||::: |||
854 snProSerValProLeuAlaValThrAlaProAspLeuThrProProVal 870
2287 CAATACCCAGCTGAAGCGGAAATTCACGGGGCAGCTCATTAATCT 2336
|||::: |||::: |||::: |||::: |||::: |||::: |||
871 GlnProThrAspLeuLeuAsnAlaAlaGlyThrSerMetThrGlyTh 887
2337 G.....ACTTGGACAGCTCTGGGATGAT..... 2361
|||::: |||::: |||::: |||::: |||::: |||::: |||
887 rAlaGluAlaGlySerThrIleThrValArgGlyProAspAsnAsnIleV 904
2362TATGACCATGGAAGCAGCTCACAAGTATATCATTCGAATAAGT 2403
::: |||::: |||::: |||::: |||::: |||::: |||
904 alGlyThrValGluValGlyGluAsnGlyGlnPheThrValProIleSer 920
2404ACAAGTATCTTGATCTCAGAGACAA 2429
|||::: |||::: |||::: |||::: |||::: |||::: |||
921 ProAlaGlnAsnAsnGlyGlnLeuLeuThrValValSerThrAspAspAl 937
2430 GTTCAATGAATCTCT.....CAAGTGAATACTACTGCT.... 2463
|||::: |||::: |||::: |||::: |||::: |||::: |||
937 aGlyAsnAlaSerValAlaValProTyrGlnThrAsnAspThrSerProp 954
2464CTCATCCCAAGGAAGCC.....AACTCTGAGGAAGTCTTTTGT 2502
|||::: |||::: |||::: |||::: |||::: |||::: |||
954 roAspLeuValThrAsnLeuAlaIleAsnAsnThrPheAsnValLeuThr 970
2503 TTTAAACAGAA.....AACATTACTTTGAAAT..... 2532
|||::: |||::: |||::: |||::: |||::: |||::: |||
971 GlyLysGlyGluAlaAsnAlaThrValThrValSerPheAspGlyAspVa 987
2533GGCACA.....GATCTTT 2545
|||::: |||::: |||::: |||::: |||::: |||::: |||
987 lIleGlyThrGlyValValAspGlyAlaGlyAsnPheGlnIleThrLeuV 1004
2546 TCATTGTCTATTTCAGGCTCTTTGATAAGGTGCGATCTG...AAATCAGAAATA 2592
|||::: |||::: |||::: |||::: |||::: |||::: |||
1004 alProAlaValGlySerLeuAlaThrLeuGluValThrGlnSerAspGly 1020
2593 TCCAACTATGACAGTATCTTTGTTTATTCCTCCACAGACTCCGCCAGA 2642
|||::: |||::: |||::: |||::: |||::: |||::: |||
1021 IleAsnThrSerAlaValAlaThrThrLeuThrProLeuIleAlaProPr 1037
2643 GACACCTAGTCTGTGTAACG 2664
|||::: |||::: |||::: |||::: |||::: |||::: |||
1037 ovalProThrAspValThr 1044
seq_name: sp_bacteria:Q9AES2
seq_documentation-block:
ID Q9AES2 PRELIMINARY; PRT; 1234 AA.
AC Q9AES2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ICA PROTEASE.
GN ICA.


```
1480 TCCATCCAGCTTGAGAGTAAGGGATTACCCCTCCAGAACACCCAGCTGGAT 1529
|||||
930 eProAla.....ProGlyHisThrAlaGly.Thr 939
1530 GAATGGCAGCATGATCGTGGACACAC.....CGTGG 1561
|||||
940 GlutPlySerAspThrAshHisTrpHisGluCysThrValAlaGly 956
1562 GAAAGGACACTTTCTTCTTATCATCCCTGGACACGCGCCTCCCAAAATC 1611
|||||
956 yCysGlyValIleIleGlu.....SerThrLysSerA 967
1612 CTCTCTCTGGATCCCACTGG.....ACAGAAGCAAGGTGCTTGT 1652
|||||
967 laHisThrAlaGlyGluTrpIleValAspThrProAlaThrAlaThr 983
1653 AGTGGACAAAACACCAAAATGGCC.....TACCTCCAAA 1687
|||||
984 AlaGlyThrLysHis.LysGluCysThrValCysHisArgValLeuGluT 1000
1688 TCCAGGCACTGCTAAGGTGGCACTTGGAAATACAGTCTGCAAGCAAGC 1737
|||||
1000 hrGlnProIleProSerThrGlyThrGluLeuLysIleIleAlaGlyAsp 1016
1738 TCACAAACCTTGACCTGACTGTCACGTCGCGTGGTCCCAATGCTACCT 1787
|||||
1017 AsnGln.....IleTyrAsnLysAlaSerGlySer..... 1026
1788 GCCTCCCAATTACAGTGACTTCCAAAGGACACAGGACACCAAAATCC 1837
|||||
1027 .....AspValThrIleThrCysAsnGlyAspPheAlaLysPhe. 1039
1838 CCAGCCCTCTGGTAGTTTATGCAAAATATTCGCCAAGGAGCCTCCCAATT 1887
|||||
1040 .....ThrGlyIleLysValAspGlySerValVal 1049
1888 CTCAGGGCCAGTGTCCAGCCCTGATTGAATCACTGAATGGAATAAAGCAGT 1937
|||||
1050 AspSerSerAsnTyrThrAla.....ValSerGlySerThrVa 1062
1938 T...ACCTTGAAGTACTGGATAATGGAGCAGGTGCTGATCTACTAAGG 1984
|||||
1062 lLeuThrLeuLys.....AlaSerTyrLeuGlyThrLeuT 1074
1985 ATGACGGTGTCTACTCAAGGTATTTTCACTATGACACCAAGTGGTAGA 2034
|||||
1074 hrAspGlySerHisThrIleThrPheValTyrThrAspGlyGluAlaAsn 1090
2035 TACAGTGTAAAGTGCGGGCTCTGGAGGAGTAAACGACGACGACGAG 2084
|||||
1091 AlaAsnLeuThrValArgThrAlaGly..... 1099
2085 AGTGATACCCAGCAGAGTGGAGCACTGTATACATACCTGGC.....TGA 2128
|||||
1100 .....SerGlyHisIleHisAspTyrGlyThrGluTrpL 1111
2129 TTGAGATGATCAATACATGGAATGGAATCCACCAACCTGAATTAAGA 2178
|||||
1111 ysSerAsnAlaAspAsnHisTrpHisGluCysAsnGlyAspLysLys 1127
2179 GATGATGTTCAACAC 2193
1128 AspGluAlaAlaHis 1132
seq_name: sp_bacteria:Q9RNI2
seq_documentation_block:
ID Q9RNI2 PRELIMINARY; PRT; 1557 AA.
AC Q9RNI2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HWA.
```

```
GN HWA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=A950006;
RA Van Schilfgaarde M., Van Ulsen P., Eijk P., Dankert J., Van Alphen L.;
RT "Characterization of adherence of non-typeable haemophilus influenzae
RL to human epithelial cells.";
DR EMBL; AF180944; RAD56660.1; -;
DR InterPro; IPR000169; Thiolprot_act_site.
DR PROSITE; PS00639; THIOLEPROTEASE_HIS; UNKNOWN.1.
SQ SEQUENCE 1557 AA; 160460 MW; 62524D384398CB37 CRC64;

alignment_scores:
.. Quality: 144.50 Length: 783
.. Ratio: 0.426 Gaps: 30
Percent Similarity: 43.295 Percent Identity: 18.263

alignment_block:
US-09-049-696-18 x Q9RNI2 ..
Align seg 1/1 to: Q9RNI2 from: 1 to: 1557

865 GACCTTTAAGAAACCACTCTATGACAAACACAGCCACCAAAATCCACCTT 914
|||
86 AspGlyAsnLysThrIleArgAsnSerValAsnAlaIleAsnTr 102

915 CTCATTGCTGCAGATTGGACAAAGAATTGTGTGTTTACTGCTTGCACAAAT 964
|||
102 pLysGlnPheAsnIleAspGlnAsnGluMetValGlnPheLeuGlnGlu 119

965 CTGGAACGATGGCCAGCTGCTAAACCCCTCAATCGA..... 999
|||
119 erSerAsnSerAlaValPheAsnArgValThrSerAspGlnIleSerGln 135

1000 .....CTGAATCAAGCAGCGCCAGCTTTCTGCTGCAG..... 1032
|||
135 LeuLysGlyIleLeuAspSerAsnGlyGlnValPheLeuIleAsnProAs 152

1033 .....ACATTGAGTGGGCTCTGGTGGTGGATGGTGACATTTGAC. 1074
|||
152 nGlyIleThrIleGlyLysAspAlaIleIleAsnThrAsnGlyPheThr 169

1075 ..ACTGCTGCCCATGTACAAAGTGAACATCATACACATAAACAGTGGCAGT 1122
|||
169 laSerThrLeuAspIleSerAsnGluAsnIleLysAlaArgAsnPheThr 185

1123 GACAGGACACACACTCGCCAAAGATTACCTGCAGCAGCTTCAGGAGGGAC 1172
|||
186 LeuGlnThrLysAspLysAlaLeuAlaGluIleValAsnHisGlyLe 202

1173 GTCCATTGCGACGGGCTTCGATCGCATTTACTGCTGATTAGGAGAAAT 1222
|||
202 u...IleThrValGlyLysAspGlySerValAsnLeuIleGlyGlyLysV 218

1223 ATCCAACATGAT.....GGATCTGAATTTGCTGCTG 1254
|||
218 alLysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeuLeu 234

1255 ACGGATGGGGAAGACACACTATAAGTGGTGGTCTTAACGAGAGTCAACA 1304
|||
235 Ala...GlyGlnLysIleThrIleSerAspIleIleAsnPro..... 247

1305 AAGTGGTGGCCATCATCCACACAGTCGCTTTGGGGCCCTCTCGACGCTAAG 1354
|||
248 .....ThrIleThrTyrSerIleAlaAlaProGluAsnGluAlaIleA 262

1355 AACTAGAGGAGCTGTCCAAATGACAGAGGTTTACAGACATATGCTTCA 1404
```

262 snLeuGlyAspIlePheAlaLysGlyGlyAsnValArgAlaA 278
1405 GATCAAGTTCAGAAAGCCCTCATT.....GATGCTTTGGGCCCT 1448
279 Asn...IleArgAsnGlnGlyLysLeuSerAlaAspSerValSerLysAs 294
1449 TTCATCAGCAAAATGGAGCTCTCTCAGCGCTCCATCCAGCTTGAGAGTA 1498
294 pHisSerGlyAsnIleValLeuSerAlaLysGluGlyGluAlaGluIleG 311
1499 AGGGA...TTAACCCCTCCAGAACAGCCAGTGGTGAATGGCAGCTGATC 1545
311 LyGlyValIleSerAlaGlnAsnGlnAlaLysGlyGlyLysLeuMet 327
1546 GTG..... 1548
328 IleThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLeuSe 344
1548 1548
344 rGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGluArgGlyGluG 361
1549GACAGCACCTGGGAAAGGACACT 1572
361 LyLysAsnGlyIleGlnLeuAlaLysLysThrSerLeuGluLysGlySer 377
1573 TTGTTTCTTATCACCCTGGACAGCGACCTCCCAATCTCTCTGG... 1620
378 ThrIleAsnValSerGlyLysGluLysGlyGlyArgAlaIleValTrpG 394
1621GATCCAGCTGGACAG..... 1635
394 YAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGlyLysAspIleA 411
1636 ..AAGCAAGTGGCTTTGAGTGGACAAACACCAAAATGGCTACCTC 1683
411 laLysThrGlyGlyPheVal.....GluThrSerGlyHisTyrLeu 424
1684 CAAATCCAGCAGCTTGTAGTGGCACTTGGAAATACAGTCTGCAAGC 1733
425 SerIleGlyAsnAspAlaAlaValGluAlaLysGluTrpLeuAspPr 441
1734 AGCTCAACAACCTTGACC..... 1752
441 oAspAsnValThrIleSerAsnGlyAsnAspAspGlnSerGlnLeuLysA 458
1752 1752
458 sAspArgGlyAspSerProAsnLysIleLeuAlaAspAsnLysHisThr 474
1752 1752
475 ValAsnAsnLysThrLeuSerThrAlaLeuAlaLysGlyIleGlyValAs 491
1753CTGACTGTCCACCTCCGTCGGGCCA 1777
491 nIleSerAlaLysLysLysValAsnValThrAlaAspIleAsnValHisA 508
1778 ATGCTACCTCCCTCCCAATTACA..... 1800
508 snGlyThrLeuThrLeuHisSerGluGlnGlyGlyValGluIleAsnGly 524
1801 ...GTGACTTCCAAACGACAGAC...ACCAGCAAAATCCCGACCCC 1844
525 AspIleThrSerGluGlnAsnGlyAsnLeuThrIleLysAlaGlySerTr 541
1845 TCTGGTGTATTCAAATATTCGCAAGGAGCTCCCAAT...CTCA 1891
541 pValAspValHisLysAsnIleThrIleGlyThrGlyPheLeuAsnIleT 558
1892 GGGCCAGTGTCCAGCCCTGATTCGAATGAGTA..... 1929
.....

seq_name: sp_bacteria:Q9YD91

seq_documentation_block:

558 hrAlaGlyGlySerValAlaPheGluLysAlaGlyGlyAspLysGlyArg 574
1930AAAACAGTTTACCTTGGAACTACTGGATAATGGAGC 1964
575 AlaAlaSerAspAlaLysIleValAlaGlnGlyValIleThrAlaGlySe 591
1965 AGGTGCTCAT.....GCTACTAAAGCATGACGCTGTCT 1996
591 rGlyGlnAspPheArgPheAsnAsnValSerLeuAsnGlyThrGlyArg 608
1967 ACTCAAGGTATTTCAACAATTTACACAGCAATGAGTACAGTGTAAAA 2046
608 LyLeuLysPheIleThrAlaLysGlyAsnLysGlyAsnPheSerAlaLys 624
2047 GTGGGGCTCTGGGAGGAGTTAACGCCAGCCAGCAGGAGAGTGATACCCCA 2096
625 Phe..... 625
2097 GCAGAGTGGACACTGTACATACCTGGCTGGATTGAG..... 2133
626 ...AspGlyValLeuAsnIleSerGlyAsnIleSerIleAsnHisThrA 641
2134 ..AATGATGAATACAA.....TGGAAAT 2154
641 laAsnAsnGlnLeuSerTyrPheHisArgGlnGlyTyrThrTyrTrpAsn 657
2155 CCACCAAGACCTGAAATTAATTAAGATGATGTTCAACAAGCAAGGTGTG 2204
658 LeuThrGlnLeuAsnValAspSerAspSer.....Se 668
2205 TTTCAGCAGAACATCC..... 2220
668 rPheSerLeuThrSerIleLysAspAlaIleLysValGlyGlyTyrAspA 685
2221TCGGGAGGC..... 2229
685 snAlaLysAspLysAsnThrGlyGlyIleGlyPheThrArgAspThr 701
2230TCATTTGT 2237
702 IlePheAsnValLysGlnGlyAlaArgValAspIleSerTyrThrLeuPr 718
2238 GGCTTCTGTATGCCAAATGCTCCCATACCTCTTCCCACCTGCC 2287
718 oIleSerProValLysAsnSerArgIleAlaAlaValAsnPheAspGlyA 735
2288 AATCACCAGCCTGAAGCGGGAATTCACGGGGCAGTCTCATTAATCTG 2337
735 snIleThr.....ValLysGlyGlyValValAsnLeu 746
2338 ACTTGACAGCTCTGGGAGTATGACCATGGACAGCTCACAAGTA 2387
747 LysPheAsnAlaLeuSerAsnAsnTyr..... 755
2388 TATCATTTCCAATAAGTACAAAGTATTTGATCTCAGACAGATTC.... 2433
755LysThrProGlyValAsnIleSerSerArgPheIleA 768
2434AATGAATCTCTCAAGTGAATACTACTGCTCTCATCCCAANGNA 2478
768 snValThrGluGlySerGlnLeuAsnIleThrGlySerMetProSerThr 784
2479 GCCAACTCTGAGGAAGTCTTTTGTAAACCAAGCAAAACATTACTTTTGA 2528
785 Thr.....LeuPheAsnValAlaAsnAspLeuIleI 795
2529 AATGGCAGACATCTTTTCATTGCTATTACGGCTGTTGATAAGGTCGAT 2577
795 eAsnAlaThrAsnSerPheValSerIleLysGluIleGlyThrAsp 811
seq_name: sp_bacteria:Q9YD91
seq_documentation_block:

Q92D91	PRELIMINARY;	PRT: 2340 AA.
Q92D91;		
DT	01-MAY-1999 (TREMBLrel. 10, Created)	
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)	
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE	CELL SURFACE ANTIGEN (SCA3).	
GN	RP451.	
OS	Rickettsia prowazekii.	
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;	
OC	Rickettsiaceae; Rickettsiae; Rickettsia.	
OX	NCBI_TaxId=782;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=MADRID E.	
RX	MEDLINE=99039499; PubMed=9823893;	
RA	Andersson S.G.E., Zomorodipour A., Andersson J.O.,	
RA	Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund	
RA	Eriksson A.-S., Winkler H.H., Kurland C.G.;	
RT	"The genome sequence of Rickettsia prowazekii and the origin	
RT	of mitochondria."	
RL	Nature 396:133-140(1998).	
DR	EMBL; AJ235271; CAA14908.1; -	
DR	InterPro: IPR003858; rompA_rOmpB.	
DR	Pfam; PF02708; rOmpA_rOmpB; 1.	
DR	Complete proteome.	
SK	SEQUENCE 2340 AA; 247862 MW; CC4070F93C165E26	CRC64;

```
alignment_scores:
  Quality: 143.50      Length: 809
  Ratio: 0.395         Gaps: 38
  Percent Similarity: 44.870      Percent Identity: 20.025

alignment_block:
  US-09-049-696-18 x Q9ZD91      ..

Align seq 1/1 to: Q9ZD91
from: 1 to: 2340
```

[illegible]

936 GACAAATCTGGAAGCATGGCGACT.....GGTAAACGCGCCTCAA 935
 937 AspPheGlyAlaAsnGlyGlyThrLeuGluPheAsnGlyAsnAsnThrty 953
 938 TCGACTGAATCAAGCA.....GCCAGCTTTTCCCTGCTGCACAGC 1036
 939 rAsnLeuAsnAlaIleValAsnGlyGlnAsnGlyIleLeuAsnAlap 970
 1037 TT.....CAGCTGGGGCTCTGGTGGTGGATG 1062
 970 heThrAsnLeuLysAlaSerAspAspThrIleGlyThr...ValLysIle 985
 1063 GTGACATTGTACAGTGTGCCCATGTACAAAGTAACTCATACAGATAAA 1112
 985 IleAsnIleGlyGlnIleGlyThrProGlnAsnPheThrIleGlnValas 1002
 1113 CAGTGGCAGTGCACAGGACACACTGCGCAAAAGATTACCTGCACAGCCT 1162
 1002 nAsn.....LysAsnLeuThrLeuValSerS 1011
 1163 CAGGAGGACGCTCCATCTGCAGCGGG..... 1188
 1011 erValAsnSerSerIleAsnPheGlyAspAlaAsnSerGlnLeuLeu 1027
 1189CTTCATCGGCATTTACTGTGATTAGGAAGAAATATCCAAC 1229
 1028 SerAlaProValAspGlnThrIleLysPheIleAsnAsnLeuAsnGluTh 1044
 1230 TCATGGATCTGMAATTGTCTGCTCAGCGGATGGGAGACACACATATAA 1279
 1044 rGlyGlyIleIleThrLeuAspSerAsnGlyAsnLeuThrIleS 1061
 1280 GTGGTGCTTTAACGAGGTCAAAAGAGTGTGCCTCATCCACACAGTC 1329
 1061 erGly.....AsnAsnGly.....Ile 1066
 1330 GCTTTGGGSCCTCTGCAGCTCAAGAACTAGAGAGGCTGTCCAAAATGAC 1379
 1067 LysLeuGly...SerLysGlyAsnGluLeuSerSerLeuAsn.....I 1080
 1380 AGGAGGTTTACAGACATATGCTTCAGAT...CAAGTTCAGAAACATGGCC 1426
 1080 eLysGlyLysValThrValThrAsnAspLeuAspIleGlnAsn..... 1094
 1427 TCATTGATGCTTTTGGGSCCTTTTCATCAGAAATGAGCTGTCTCTCAG 1476
 1095IleHisGlnLeuAsnIleAsnAsnGlyAlaLeu..... 1105
 1477 CGCTCCATTCAGCTTGAGAGTAAGGATTAACCTCCAGAACAGCCAGTG 1526
 1106PheAspAspGlnSerLeuThrSerAlaLysIleLysAs 1118
 1527 GATGAAT...GGCACATGATCGTGGACAGCAGCGTGGGAAAGACACTT 1573
 1118 nIleAsnIleGlyThrValAlaGlyGlyAlaThrTyThrLeuAspAlaI 1135
 1574 TCTTTCCTTATCACCTGGACACAGCGCCTCCCCAAATCCTTCTCTGGAT 1623
 1135 le..... 1135
 1624 CCCAGTGACAGAACGAAAGTGGCTTTGTAGTGGCAAAAACACCAAA.. 1671
 1136AsnAspAsnPhe.....AspLeuAsnThrSerGI 1145
 1672 ATGCCTTACCTCCAAATCCACGACATGTGTAAAGTTGGCATTTGGAAAT 1720
 1145 yMetValPheLysHisGlnAspSerIleLeuLeuLeu.....LysA 1159
 1721 ACAGTCTGCAAGCAAGCTCACAAACCTTGACCTTGACTGTCAAGTCCCGT 1770
 1159 snSerSerAsnThrAsnAspHisThrIleThrLeuThrSerAlaLeuAsp 1175

Tue Apr 2 09:39:52 2002

```
1771 GGGTCAATGCTACCTGCTCCCAATTACAGTGACTTCCAAACGACAA 1820
      |||
1776 ProGlyAsnAsnGlnPheGlyLeuLeuLeuLeuLeuLeuLeuLeu 1192
      |||
1821 G.....CACACCAGCA 1831
      |||
1822 LeuThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 1209
      |||
1832 AATTCCTCCAGCCTCTGCTAGTTATGCAATATTCGCAAGGAGCCTCC 1881
      |||
1833 sHisMetLeuLeuGlnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 1224
      |||
1882 CCAATTCCTCAGGCGCCAGTGTCTACACCTCTGATGTAATCAGTGA 1931
      |||
1883 ..LeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 1234
      |||
1932 ACAGTTACTTGGAACTACTGAT 1956
      |||
1933 uAsnValThrLeuAsnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 1251
      |||
1957 .....AAT 1959
      |||
1958 snValLeuPheAsnLysAsnThrThrThrAlaThrGlyAsnLeuLeu 1267
      |||
1960 GGA.....GCAGTGTCTGCTACTAAGATGA 1988
      |||
1961 GlyHisValAspPheGlnGlyAsnAlaGlyValLeuLeuLeuLeu 1284
      |||
1989 C...GGTGTCTACTCAAGGTATTTCACAACTTATGACACGAATGSP 2035
      |||
1990 pLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 1299
      |||
2036 ACAGTGTAAAGTGGCGCTCTGGAGGAGTTAACCCAGCCAGCAGCA 2085
      |||
1300 ..ThrLeuAsnPheAsnGlySerGlyLysValThrGlyLeuLeuLeu 1315
      |||
2086 GTGATACCCAGCAGAGTGAGCA.....CTGTACATACCTG 2123
      |||
1316 IleValMetLeuGlnAlaGlyAlaGlyAspValSerLeuSerAla 1332
      |||
2124 CTGATTCAGATGATGAATACATGATGATCCACCAAGA..... 2163
      |||
1332 yAsnTySerIleThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1349
      |||
2164 .....CCTGAAATTAATAAGATGATTTCAA 2190
      |||
1349 heAlaAlaAsnSerHisLeuThrThrAspLeuLeuLeuLeuLeu 1365
      |||
2191 CACAAGCAAGTGTGTTTCAGCAGAACATCTCTGGAGGCTCATTTGG 2240
      |||
1366 AspLeuAsnLeuValPhe.....IleAsnGlyGlySerValSer 1379
      |||
2241 TTCATGATCCCAATGCTCCATACCTGATCTCTCCACCTGGCCAAA 2290
      |||
1379 ySerIleGlyAlaAsnAlaAlaValGlyAspIleIleLeuLeuLeu 1396
      |||
2291 TCACGACCTGAAGCGGAAATTCACGGGCGAGTCTCATT..... 2331
      |||
1396 erValAsnPheSerAsnThrLeuLysSerGlyAsnIleValIleSer 1412
      |||
2332 .....AATCTGACTTGGACA.....GCTCCG 2354
      |||
1413 GlyAlaThrMetGlnValAsnAsnValThrAlaThrAspIleSer 1429
      |||
2355 GGATGATATGACCATGGAACAGCTCACAGTATATCATTCGAATA 2404
      |||
1429 yLysAsnAlaAsnAsnGlyThrLeu..... 1437
      |||
2405 CAAGTATTCATTCAGACAGACAGTTCATCACTCTTCAGTGAAT 2454
      |||
1438 .....LysLeuAsnAsnHisThrProIleAsn 1446
      |||
2455 ACTACTGCTCTCATCCCAAGGAA.....GC 2480
```

```
1147 IleThrSerThrLeuGlyAsnAsnAlaIleGlyThrIleGluValAl 1463
      |||
2481 CAACTCTGAGGAAGTCTTT.....TTGTTTAAACCAGAAACATTA 2524
      |||
1463 aAsnAsnAspValThrIleThrGlyThrLeuGlnAlaGlnAsnIle 1480
      |||
2525 TTGAAATGCGACAGATCTTTTCATTGCTATTGAGGCTGTGATAAG 2574
      |||
1480 heSerAsnAlaThrGln.....AlaAlaThrLeuThrLeu 1491
      |||
2575 GATCTGAAATCAGAAATATCCACATT 2601
      |||
1492 GlyAlaAlaSerGlnValThrAsnIle 1500
      |||
seq_name: sp_bacteria:P71401
seq_documentation_block:
ID P71401 PRELIMINARY; PRT: 2353 AA.
AC P71401;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HSF PROTEIN.
GN HSF.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=C54;
RX MEDLINE=97047989; PubMed=892830;
PA Gene J.W., Cutter D. III, Barenham S.J.;
RT "Characterization of the genetic locus encoding Haemophilus influenzae
   type b surface fibrils."
RL J. Bacteriol. 178:6281-6287(1996).
DR EMBL; U41852; AAC44560.1;
SQ SEQUENCE 2353 AA; 243844 MW; 53E491B51A897B77 CRC64;
```

alignment_scores:
Quality: 141.50 Length: 967
Ratio: 0.321 Gaps: 44
Percent Similarity: 45.605 Percent Identity: 18.718

alignment_block:

US-09-049-696-18 x P71401 ..

Align seg 1/1 to: P71401 from: 1 to: 2353

```
274 ACATGGAACACAAAGCTGACTATGTGACACCAAACTTGAGACCTACAA 323
      |||
429 SerTrpLysAlaLysAlaGlu..... 435
324 AATGCTGATGTTGTTGTTGCTGAGTCTACTCTCCAGGTAATGATGAC 373
      |||
436 .....AlaAsp.....ThrAspGlyAlaLeuGluGlyI 445
      |||
374 CTCTACACTGAGCAGATGGCACTGTGGAGAGAGGGTGAAGAGTCCAC 423
      |||
445 leSerLysAspGlnGluValLysAlaGlyGlu..... 455
424 CTCACTCTCTGATTTTCATTTCAGGAAAAAGTTAGCTGAATGGACCACA 473
      |||
456 ...ThrValThrPheLysAlaGlyLysAsnLeu...LysValLysGlnAs 470
      |||
474 AGGTAGGCAATTGTCATGAG.....TGGGCTC 502
      |||
470 pGlyAlaAsnPheThrThrSerLeuGlnAspAlaLeuThrGlyLeuThrS 487
      |||
503 ATCTACGATGGGAGTATTTGACGAGTACATAATGATGAGAAATCTTAC 552
      |||
```

487 erlleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspAlaLysThrVal 503
553 TTATCC...AATGAGAATACAAAGTAGTAGTTCAGCAGCGTATTAC 599
504 IleAsnLysAspGlyLeuThrIleThrProAlaGlyAsnGlyGlyThrTh 520
600 TGGTACAAATGTAGTAAAGAGTGTCTAGGAGGCGAGCTGTGTACACAAA 649
520 rGlyThrAsnThrIleSerValThrLysAspGlyIleLysAlaGlyAsnL 537
650 GATGCACATTCAATAAGTACAGGATC...TATGAAAAGGATGT 693
537 ysAlaIleThrAsnValAlaSerGlyLeuArgAlaTyAspAspAlaAsn 553
694 GAGTTTCTTCCAAATCCCGCAGCAGGAGAAGCTTCTATATGTTTGC 743
554 PheAspValLeuAsnAsnSerAlaThrAsp...LeuAs 565
744 ACAACATGTTGATCTTATAGTGAATTTCTGTACAGAACAAAACCAACA 793
565 nArgHisValGluAspAlaTyLysGlyLeuLeu...AsnLeuAsnG 580
794 AAGNAGCTCCAAACACCAAAATCAAAATCCAATCTCCGAGCACATGG 843
580 LuLysAsnAlaAsnLysGln...ProLeuValTh 590
844 GAAGTGATCCGTGATTCTGAGGACTTTAAGAAAACCACTCTATGACAAC 893
587ProLeuValTh 590
894 ACAGCCACCAAAATCCACCTTCTCATCTGCTGAGATTGGACAAAGATTG 943
590 rAspSerThrAlaAlaThrValGlyAspLeu...ArgLysLeuG 604
944 TGTCTTAGTCTTGCAATCTGGAGCATGGCGACTGTGTAACCGCTC 993
604 lyTrpValValSerThrLysAsnGly...ThrLysGluGluSer 617
994 AATCGACTGAATCAAGCAGCGCAGCTTTCTGCTGCGCAGCAGTGTGAGCT 1043
618 AsnGlnValLysGlnAlaAspGluVal...Le 627
1044 GGGTCTCTGGGTGGGATGGTGTGACATTGTGACAGTGTGCCATGTACAAA 1093
627 uPheThrGlyAlaGlyAlaAlaThrValThrSerLysSerGluAsnGlyL 644
1094 GTGAACTCATACATAAACAGTGGC... 1119
644 ysHisThrIleThrValSerValAlaGluThrLysAlaAspCysGlyLeu 660
1120 ...AGTGACAGGGACACACTGCCAAAAGATTACCT... 1152
661 GluLysAspGlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAs 677
1153GCACGACTTCAGGAGGAGCTCCATCTGCGCGGGCTC 1192
677 pAsnValLeuThrValGlyAsnAsnGlyThrAlaValThrLysGly... 692
1193 GATCGGATTTACTGTGATTAGGAGAATATCCAACTGATGGATCTGAA 1242
693GlyPheGluThrVal...LysThrGlyAlaThrAspAlaAspArg 706
1243 ATTGTGCTGTGCGGATGGGAGAGACACACTATAGTGGTCTTTAA 1292
707GlyLysValThrValLysAspAlaThrAlaAs 717
1293 CGAGGTCAAAACAAAGTGGTGCATCCACACAGCTGCTTTGGGGCCT 1342
717 nAspAlaAspLysValAlaThrValLysAspValAlaThrAlaIleA 734
1343 CTGAGCTCAAGAACTAGAGAGCTGTCCAAAATGACAGGAGGTTTACAG 1392
734 snSerAlaAlaThrPheValLysThrGluAsnLeuThrThrSerIleAsp 750

1395 ACATATGCTTCAGATCAAGTTTCAGAACAAATGGCTCATTTGATGCTTTGG 1442
751GluAspAsnProThrAspAsnGlyLysAspAsp... 761
1443 GGCCTTTTCATCAGGAAT...GGAGCTCTCTCTCAGCGCT 1480
762 AlaLeuLysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuL 778
1481 CCATCCAGCTTGAGAGTAAGGATTACCTC...CAGAACAGC 1521
778 ysValLysArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeu 794
1522 CAGTCGATGAATGGCAGATGATGCTGCACAGCACCGTGGGAAAGGACAC 1571
795 GluValLysThrAlaLysValSerAspThrLeuThrIleGlyGlyAsnTh 811
1572 TTTGTTTCTTATCACCTGGCAGCAGCAGCTCCCAATC... 1611
811 r.....ProThrGlyGlyThrThrAlaThrProLysValAsnIleThrS 826
1611 1611
826 erThrAlaAspGlyLeuAsnPheAlaLysGluThrAlaAspAlaSerGly 842
1612CTTCTCTGGATCCAG 1628
843 SerLysAsnValTyLeuLysGlyIleAlaThrThrLeuThrGluProSe 859
1629 TGGACAGAACGAAGTGGCTTTGTA.....CTGGACAAAAACACA 1669
859 rAlaGlyAlaLysSerSerHisValAspLeuAsnValAspAlaThrLysL 876
1670 AAATGGCTTACTCCAAATCCAGCATTTGTAAGTTGGCAGCTGGAAA 1719
875 ysSerAsnAlaAlaSerIleGluAspValLeuArgAlaGly... 889
1720 TACAGTCTGCAAGCAAGCTCACAACTTGACCTGCTGCTGCTCCCG 1769
890 TrpAsnIleGlnGlyAsnGlyAsnValAspTyThrValAlaThrTyraS 906
1770 TGCCTCCAAATGCTACC.....CTGCCTCCAAATACAGTGA 1804
906 pThrValAsnPheThrAspSerThrGlyThrThrThrValThrValT 923
1805 CTTCCAAACGAACAGCAGCACACCAATTCGCCAGCCCTCTGTAGTT 1854
923 hrGlnLysAlaAspGlyLysGly... 930
1855 TATCAATATTCGCCAAGGAGCTCCCAATTTCTCAGGGCAGTGTAC 1904
931 ...AlaAspValLysIleGlyAlaLys...Th 939
1905 AGCCTGATTAATCAGTGAATGGAACACAGTTACC...TTGGAACATAC 1951
939 rSerValIleLysAspHisAsnGlyLysLeuPheThrGlyLysAspLeuL 956
1952 TGGATAATGGAGCAGGCTGCTGCTACTACTAGGATGACGGT... 1992
956 ysAspAlaAsnAsnGlyAlaThrValSerGluAspGlyLysAspThr 972
1993GTCTACTCAAGGTATTTCACAACTTATGACACCAATGG 2030
973 GlyThrGlyLeuValThrAlaLys...ThrValIleAspAlaValAs 987
2031 TAGATACAGTGTAAAGTGGCGGCTCTGGGA... 2061
987 nLysSerGlyTrpArgValThrGlyGluGlyAlaThrAlaGluThrGlyA 1004
2062GGAGTTAACGCCAGCAGGAGAGTATACCCAGCAGAGTGA 2106
1004 laThrAlaValAsnAlaGlyAsnAlaGluThrValThrSerGlyThrSer 1020

```

2107 GCACGTGATACATCTGGCTGGATTCAGAAATGATGAATACAAATGGAATCC 2156
    : : : : :
1021 ValAsnPhLeAsnGly.....AsnAl 1028
    : : : : :
2157 ACCAAGACCTGAAATATATAAGGAT.....G 2182
    : : : : :
1028 aThrAlaThrValSerLysAspAsnGlyAsnValLysTyrA 1045
    : : : : :
2183 ATGTT.....CAACACAAGCAAGTGTGT 2205
    : : : : :
1045 spValAsnValGlyAspGlyLeuLysIleGlyAspLysLysIleVal 1061
    : : : : :
2206 TTCACAGAACATCC.....TCGGGAGCTCATTTGTG..... 2238
    : : : : :
1062 AlaAspThrThrLeuThrValThrGlyGlyLysValSerValProAl 1078
    : : : : :
2239 ....GCTCTGATGCTCCAAATGCTCCCATACCTGATCTCTCCACCTG 2284
    : : : : :
1078 aGlyAlaAsnSerValAsn..... 1085
    : : : : :
2285 GCCAAATACCGACCTGAAGCGGGAATTCACGGGGCAGTCTCATTAAT 2334
    : : : : :
1086 ..AsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsn 1101
    : : : : :
2335 CTGACCTGACACCTCTCTGGGATGATTATGACCATGGAAACAGCTCAAA 2384
    : : : : :
1102 LeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyGluSerGlu 1118
    : : : : :
2385 GTATATC..... 2391
    : : : : :
1118 yGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysAla 1135
    : : : : :
2392 .....ATTGGAATAGTACAGTATTCCTGATCTCAGACAAAGTTC 2433
    : : : : :
1135 LysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSerLeu 1151
    : : : : :
2434 AATGAATCTTCAAGTGAATATCTGCTCTCATCTCCCAAGAACCCAA 2483
    : : : : :
1152 GlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAla 1168
    : : : : :
2484 C.....TCTGAGGAAGCTTTTGTGTTAAACCAAGAAACATTAATTTT 2526
    : : : : :
1168 nGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrIle 1185
    : : : : :
2527 .....GAAATGGCAGCAT 2541
    : : : : :
1185 hrLeuAlaAsnGlyAlaAlaAlaGlyThrAspAlaSerAsnGlyAsnThr 1201
    : : : : :
2542 CTTTTCATTCATTCAGGCTCTGTATAA..GGTCGATCTGAAATCAGAAA 2590
    : : : : :
1202 TleSerValThrLysAspGlyIleSerAlaGlyAsnLysGluIleThrAs 1218
    : : : : :
2591 TATCAACATTCACAGATATCTTTGTTATTCCTCCACAGACTCCGCCA 2640
    : : : : :
1218 nVallys.....SerAlaLeuLysThrTyrLysAspThrGlnAsnThrA 1233
    : : : : :
2641 GAGACACCTAGTCTCTGATGAACGCTGCTCTGCTCTTAATATTCATAT 2690
    : : : : :
1233 laAspGlu..... 1235
    : : : : :
2691 CAACAGACCATTCCTGCAATTCACATTTTAAAAATATTGTGGAAGTGA 2740
    : : : : :
1236 ThrGlnAspLysGluPheHisAlaAlaValLysAsnAlaAsnGluVal 1252
    : : : : :
2741 T 2741
    : : : : :
1252 u 1252
    : : : : :
seq_name: sp_invertebrate:Q9U7P4
seq_documentation_block:
ID Q9U7P4 PRELIMINARY; PRT; 494 AA.
AC Q9U7P4;

```

```

01-MAY-2000 (TRENBLrel. 13, Created)
01-MAY-2000 (TRENBLrel. 13, Last sequence update)
01-JUN-2001 (TRENBLrel. 17, Last annotation update)
HYPOTHETICAL 55.0 KDA PROTEIN (FRAGMENT).
Eufolliculina uhligi.
OC Eukaryota; Alveolata; Ciliophora; Spirotrichea; Heterotrichida;
OC Eufolliculina.
OX NCBI_TaxID=46026;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99444925; PubMed=10517337;
RA Markmann-Mulisch U., Reiss B., Mulisch M.;
RT "Cell type-specific gene expression in the cell cycle of the dimorphic
RT ciliate Eufolliculina uhligi.";
RL Mcl. Gen. Genet. 262:390-399(1999).
DR ENBL; AF121336; AAF13350.1; -.
DR InterPro: IPR002035; WFA.
DR InterPro: IPR003015; HLH_Myc.
DR Pfam; PF00092; vwa; 1.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
DR PROSITE; PS0234; WFA; 1.
DR SMART; SM00327; VWA; 1.
KW Hypothetical protein.
FT NCN_TER 1
SQ SEQUENCE 494 AA; 54960 MW; 4A69C8B892D7B24 CRC64;

alignment_scores:
    Quality: 140.50      Length: 420
    Ratio: 0.614        Gaps: 19
    Percent Similarity: 54.524      Percent Identity: 22.381

alignment_block:
US-09-049-696-18 x Q9U7P4 ..
Align seg 1/1 to: Q9U7P4 from: 1 to: 494

895 CAGCCACCAAAATCCACCTTCTTCATTTGCTGCAGATTGCGACAAAGAATTGT 944
    : : : : :
76 GluSerProAlaGlnThrSerGluAlaSerArgSerGlyValAspIleVal 92

945 GGTGTTAGTCTTTCACAAATCTGGAAGCATGCGGACTGGTAACCCCTCA 994
    : : : : :
92 ICys...ValIleAspValSerGlySerMet...GlnGlyGluLysIle. 106

995 ATCGACTGAATCAAGCAGCCAGCTTTCTCTGTCGACAGCTTGAGCTG 1044
    : : : : :
107 ..GlnLeuValGlnThrThrLeuAsnPheMetValGluArgLeuSerPro 122

1045 GGTCCTCGGTTGGGATGGTGCATTTGACATTTGACATGCTGCCCATGTACAAAG 1094
    : : : : :
123 AlaAspArgIleCysLeuIleSerPheSerAsnAspAlaThrLysIleSe 139

1095 TGAACATCATACATAAACAGTGGGAGTCAGACAGGACACACACTCGCAAAA 1144
    : : : : :
139 rArgLeuValGlnMet...SerProLysGlyLysLysGlnLeuLysSerM 155

1145 GATTACT...GCAGCAGCTTCAGGAGGACGCTCCATCTGCAGCGGCTT 1191
    : : : : :
155 etileProArgLeuValAlaSerGlyThrAsnIleValGlyGlyLeu 171

1192 CGATCGGCATTTACTGTGATAGGAAGAATATCCAACTGATGGA...TC 1238
    : : : : :
172 GlutyrGlyLeuGlnAlaLeuArgGlnArgThrIleAsnGlnLeuSe 188

1239 TGAATTTGCTGCTGACGATGGGGAAGACAACTATAAGTGGTGTCT 1288
    : : : : :
188 rSerIleIleLeuLeuSerAspGlyGlnAspAsnGlyThrThrVal 205

1289 TTAACGAGGTCAA...CAAGTGGTGGC 1314
    : : : : :
205 euGlnArgAlaLysAlaThrMetAspSerIleValIleArgAspTyr 221

```



```
1171 ACGTCCATCTGCAGCGCGCTTCGATCGCATTT.....ACTGTGATTAG 1214
531 .....valProAsnValIysGlySerPheGlnGluValValIe 544
1215 GAAGAAATATCAACTGATGATCTGAAATGTGTGCTGACGATGGG 1264
544 tlyLys.....IleTyrLeuThrGlnGlyG 553
1265 AGACAACACTATAGTGGTCTTACAGGAGTCAACAAGTGGTCC 1314
553 lu....HisThrLeuSer.....LeuGlnLysSerGlyGly 563
1315 ATCATCCACACAGTGCCTTTGGGCGCTCTGCAGCTCAAGAACTAGAGGA 1364
564 .....TyrThrIleAlaIleAspTyrPheThrIleGluGluLeuVal 578
1365 GGTGTCAAAATGACAGGAGTTTACAGACA..... 1395
578 uAlaAsnLysAsnLysIleSerValAspThrLysLeuValThrProAsn 595
1396 ..TATGCTTCAGATCAAGTTCAGAAATGGCTCATTGATGCTTT... 1440
595 roHisProAsnAlaGlnLysLeuMetLysTyrLeuAlaSerIleTyrGly 611
1441 .....GGGCGCCTTCATCAGAAATGGAGTGTCTCTCA 1475
612 GluLysIleLeuSerGlyGlnGlnSerGlyAspGlyLysGluIleG 628
1476 GCGCTCCATCCAGTGCAGAGTAAGGATTACCCCTCCAGACAGCCAGT 1525
628 nMetIlePheAspValThrLysArgTyrProAlaValArgGlyPheAsp 645
1526 GGATG.....AATGGCAGAGTGCCTGGAGCAGCAGCCGTGGAAAGGAC 1569
645 heMetAspTyrSerProSerArgValGluHisGlyThrLysGlyThrAsp 661
1570 ACTTTGTTTCTTATCCTACCTGCAATCCAGGCAATTCAGTGGTGGC 1619
662 ValGluGluAlaIleLysTrp..... 669
1620 GGATCCAGTGGACAGACAGAGTGGCTTTGTAGTG.....GACA 1660
669 p.....LysSerGlyGlyIleValAlaPheCysTyrPheHst 681
1661 AAACACCAAAATGGCTACCTCCAAATCCAGGCAATTCAGTGGTGGC 1710
681 rAsnAlaProThrGlyLeuIleAspGlnProGly.....Lys 693
1711 ACTTGGAAATACAGTCTGCAAGCAAGCTCACAAACCTTGACCTGACTGT 1760
694 GluTyrTrpArgGlyPheTyrThrGluAlaThrThrPheAspIleLys 710
1761 CAGTCCCGTGGCTCAATGCTACC.....CTGCCTCCAA 1795
710 sAlaMetAspAsnProAsnSerThrGluTyrLysLeuIleLeuArgAsp 727
1796 TTACAGTACTTCCAAACCAAGCAAGCAGACCAATTCGCCGCCCT 1845
727 leAspAlaIleAlaGluGlnLeuLysArgLeuGlnLysGluGlyValPro 743
1846 CTGCTAGTTTATGCAATATTCGCCAA..... 1872
744 ...ValLeuPheArgProLeuHisGluAlaSerGlyGlyTrpPheTrp 759
1873 .....GGAGCTCCCAATTCAGGGCC..... 1896
759 pGlyAlaLysGlyProGluProTyrIleLysLeuTrpLysLeuMetPhe 776
1897 .....AGTGTACACCCCTGATTGAATCAGTCAAT 1926
776 spArgLeuValAsnTyrHisLysLeuAsnAsnLeuIleTrpValTrpAsn 792
1927 GGAACACAGTTACCTTGGAACTACTGGATAATGGAGCA.....GG 1967
```

```
||||| ..... ||||| |||
798 GlyGlnAspAlaAlaTrpTyrProGlyAspAsnTyrValAspIleIleG 809
1968 TGTGATGCTACTAGGATGACGGTGTCTACTCAAGGTATTTCACAACCT 2017
809 yGluAspIleTyrGluGluLysAlaGlnTyrSerProTyrAlaAspArg 826
2018 ATGACAGCAATGGTAGATACAGTGTAAAAGTGGCGGCTCTGGAGGAGTT 2067
826 he.....AlaLysAlaLeuLysTyrThr 833
2068 AACGACGCCACGAGAGTGTATACCCAGCAGAGTGGAGCAGTGTACAT 2117
834 Asn...AlaArgLysMetIleAlaLeuThrGluCysGlyThrIleProAs 849
2118 ACCTGGCTGGATTGAGATGATGAATACAATGAAT 2154
849 pProAlaValLeuLysGlnGluGlyIleSerTrpSer 861
seq_name: sp_bacteria:Q9PHN8
seq_documentation_block:
ID Q9PHN8 PRELIMINARY; PRT; 978 AA.
AC Q9PHN8;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-MAR-2001 (Tremblrel. 15, Last sequence update)
DE POSSIBLE LIPOPROTEIN.
GN CJ0629.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RA MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
BA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
FA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
QA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrall B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139075; CAB/5265.1; -.
KW Complete proteome.
SQ SEQUENCE 978 AA; 102498 MW; 06DAD1A5FA433609 CRC64;

alignment_scores:
Quality: 140.00 Length: 1010
Ratio: 0.328 Gaps: 46
Percent Similarity: 42.277 Percent Identity: 19.010
alignment_block:
US-09-049-696-18 x Q9PHN8 ..
Align Seg 1/1 to: Q9PHN8 from: 1 to: 978
7 ACAGGAGATGTACAGCAATGGGCAATTAAGATTCGTGTGTTCATCTT 56
||||| ..... ||||| |||
25 ThrGlyGluSerIleTyrPheGlyAsnAlaLysIleSerSerPheValAs 42
57 G.....ATCTCTACCTTCTAGAGGGGCCCTGAGTAAT..... 90
42 nSerGlyThrIleLysSerLysGlnGlyAlaGlyValAsnIleSerGlnG 59
91 .....TCACTCATTCAGTCAGACAAATGGCTATGAAGGCAATTCGCT 135
||||| ..... ||||| |||
59 lyThrSerIleGluAsnPheAsnAsnThrGly...ThrGlyIleIleGlu 74
135 GCAATCGACCCCAATGTGCCAGAGATGAACACTCAATTCACAACAAATAA 185
```

75 GlyLysArgMetGlyValAsnValArgSerThrIleAsnThrPheValAs 91
186 GGACATGGTGACCGACATCTCTGTATGTTGTAAGCTACAGGAAGC 235
91 nAsp.....GlyLeuIleAlaAlaThrAsnAspG 101
236 GATTTTATTCAAA...AATGTGCCATTTTGTATCTCTGAA...ACATGG 279
101 lylleGlnIleAsnAlaAsnValLysThrLeuIleAsnLysGlyThrIle 117
280 AAGACAAAGGCTGACTATGTGAGACCA.....AACTTGAGACCTA 320
118 LysGlyAspAlaIleSerIleArgSerLeuGlyGlyThrIleGluThrLe 134
321 CAAAAATGCTGATGTTCTGTTGCTGAGTCTACTCTCCAGGTAATGATG 370
134 uThrAsnGluGlyIleMetThrGlyLysSerAlaGlyIleThrMetAsnA 151
371 AACCTACACTGACCATGGGCAACTGTGGAGAGAAGGTGAAGGATC 420
151 rgSerLeuValLysThrLeuThrAsnSerGlyThrIleAsnGlnAsnAsn 167
421 CACCTCACTCTGATTTCTTTCAGGAAAAAGTTAGCTGAATATGGACC 470
168 SerAlaThr...TrpSerAlaGlyIleLysLeu..... 177
471 ACAAGGTAGGCAATTTCTCCATGATGGGCTCATCTACGATGGGAGTAT 520
177 177
521 TTGACGAGTACAATAATGATGAGAAATCTACTTATCCATGGAGAATA 570
178GluAsnGlySerIleleGluAsnIleIleAsnThrGlySerIle 192
571 CAAGCA.....GTAAGATGTTACAGAGCTATTACTGTGTACAAA 608
193 ArgSerAsnAlaPheGlyIleSerValThrGlyGlyLysPheGlyThrLe 209
609 TGAGTAAAGAGGTGTCAGGAGGACGCTGTTACACCAAA..... 648
209 uThrIleLys.....AspGlyGlyMetValThrGlyLysThrSerAlaI 224
649AGA 651
224 leGlyValGlyArgSerGlnThrLeuGlyAspLeuThrIleAspGlyArg 240
652 TGACATTCATTAAGTAAGTAACAGGACTCTATGAAAAGGATGTGATTTGT 701
241 SerAsnAsnGlyThrValSerGlyIleThrSerGluGluHisGlyIleLe 257
702 TCTCCAAATCCCGCCAGACGAGGAGGCTTCTATATGTTTGCACACATG 751
257 uLeuGluAsn..... 260
752 TTGATTCATAGTTGAATTCGTACAGACAAAACCAACACAAAGAGCT 801
260 260
802 CCAACAAAGCAAAATCAAAATGCAATCTCCGAGACACATGGGAAGTAT 851
261 ...AsnSerArgThrGlnLysIleGluLeuLysAsn...GlyGlyIleI 275
852 CCGTGAATCTGAGACATTTAAGAAAACCACTCTATGACAACACAGCCAC 901
275 eLysGlyAsnIleAsp..... 280
902 CAAATCCCACTTCTCTCTGCTGAGATTCGACAAACAAATTTGTGT... 948
281GlyIleArgLeuIleAsnSer 287
949TTAGTCTTGACAAATCTGGAAGCATGGCGAC 980
..... 999

288 AlaSerLeuSerGlyGluMetIleLeuSerGlyGluGlySerArgValG 304
981 TGTAAACCGC...CTCAATCGACTGAATCAAGCAGGCGCAGCTTTTCCTGC 1027
304 uGlyGlyArgGlyValGlyIleLeuAsnArgSerGlyLysIleGluGly 321
1028 TGCACACAGTTGACTGGGCTCTCGGTT..... 1056
321 erIleLysValGluAspGlyAlaThrValThrAlaThrSerAsnArgAla 337
1057GGGATGGTGACA..... 1068
338 IleAlaAsnSerGlySerGlySerIleThrGlyGlyIleThrValSerG 354
1069 1068
354 yLysAsnThrLysLeuGluGlyAsnIleAsnThrGlyAsnAlaSerI 371
1069TTTGACAGTCTGCCCATGTACAAAGTGAATC 1101
371 leGlySerAspIleLysIleGlyGlyAlaLysValGluGlyLeu 387
1102ATACAGATAAACAGTGG 1118
388 ValAsnGlnGlyAsnGlySerIleSerGlySerValGlnValSerGly 404
1119 CAGTGACAGGACACACTCGCCAAAGATTACCTGCACAGCTTCAGGAG 1168
404 ySerIleAspSerIleThrAsnGluGlyAsnGlyAlaIleSerGly 421
1169 GGAGCTCCATCTGC.....AGCGGGCTTCGATCGGCACTTACTGTGATT 1212
421 erIleThrValThrLysAspSerLysLeuAspSerIleThrAsnThrSer 437
1213 AGGAAGAAATATCAACTGATGATCGAAATGTGTCTGCACGAGTGG 1262
438 ThrSerSerThrGlyIleSerGlySer.....IleThrAsnAs 450
1263 GGAAGACAACTATAAGTGGTGCTTTAACGAGGTCAAAACAAAGTGTG 1312
450 nSerAspAsnLysLeu.....GluIleSerAsnSerGlyA 462
1313 CCATCATCCACACAGTCTGCTTGGGGCCTCTGCAGCTCAAGAACTAGAG 1362
462 snIle..... 463
1363 GAGCTGTCCAAAATGACAGGAGGTTTA.....CAGACATATGCTTCAGA 1406
464GlyGlyLysIleGluSerThrGlySerAlaAs 474
1407 TCAAGTT.....CAGAAATGGCCTCATTTGATGCTTTTGGGGGCC 1447
474 pMetValIleSerAsnSerAsnGlyGlyThrIleSerGly...GlyIle 490
1448 TTTTCATCAGGAATGGAGCTGCTCTCAGCGCTCCATCCAGCTTGAGAT 1497
490 erSerSerGlySerGlySerThrSerIleSerAsn.....Ser 502
1498 AAGGGATTAACTCCACAGACAGCAGTGGATGAATGGCAGCTGATCCT 1547
503 GlnGlySerThrIleAsnAsnGlyIleThrValSerGlySerAlaGlnVa 519
1548 GGAC.....AGCACCGTGGGAAGGACACATTTGTTCTTATCA 1585
519 lGluIleSerAsnGlnGlySerValGlyLysAsp..... 530
1585 CCTGGACAACCCAGCTTCCCAAAATCTCTCTGGGATCCCGAGTGGACAG 1635
530 530
1636 AAGCAAGTGGCTTTGTAGTGGACAAAACACAAATGGCTTACCTCCA 1685
532 ...GluAsnGlyAsnThrValThrAsnAsn..... 539

```
1686 ATCCAGGCTTCTAAGGTTGGC.....ACTTGAATACAGCTCTG 1729
      |||:::
540 .....GlySerGlySerValGlyIleLysAspThrLeuValSerThrA 554
      |||:::
1730 AAGCAAGCTCACAAACCTTGACCCCTGACTGTACAG.....TCGGTGGC 1773
      ::::: |||:::
554 spLysAsnThrGlyLysLeuAsnThrValValIleGlyGlySerArgAla 570
      |||:::
1774 TCATGTCTACCTCGCTCCCAATTCAGTACTGCTCCAAACGACAAAGA 1823
      |||::: |||:::
571 PheAsnValLysValGluAsnIleThrVal...AspGlnSerAsnValAs 586
      |||:::
1824 CACCAGCAAAATCCCGAGCCCTCTGGTAGTTATGCAAAATATTCGCCAAG 1873
      |||:::
586 pleuGluGluLeu..... 590
1874 GAGCCTCCCAATCTCAGGGCCAGTGTACAGCCCTGATTCGAATCAGTG 1923
      |||::: |||:::
591 .....AsnAspIleAsnAsnIleIleSerGlyVal 600
      |||:::
1924 AATGGAANAACAGTTACCTTGGAACTACTGGATAATGGAGCAGGTCTGA 1973
      |||::: |||:::
601 AsnGlnAsnAsnIleGly...AsnIleGlyThrAsnGlySerGlyClnI 616
      |||:::
1974 TGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCACAACTTATGACA 2023
      |||::: |||:::
616 eSerLeuSerPheAspProIleThrGlyLysLeuThrThrAspPheAsnL 633
      |||:::
2024 CGAATGCTAGATACACT...GTAAAAGTCGGGCTCTGGAGAGCTTAAC 2070
      |||::: |||:::
633 euAsnAlaSerIleSerGlyAlaThrPheArgSerLeuIleSerThrThr 649
      |||:::
2071 GCAGCCAGCGGAGAGTGATATACCCAGCAGAGTGGAGCAGCTGTACATACC 2120
      |||:::
650 SerArgArgSerThrPheIle..... 656
2121 TGCGTGGATTGAGAAATGAAATACAAATGGAATCCACCAAGACCTGAA 2170
      |||::: |||:::
657 .....AspAsnValMetGlyAsn..... 662
2171 TTAATAAGGATGATGTTCACAAACAGCAGTGTTCACAGCAACATCC 2220
      |||::: |||:::
663 .....SerMetGlnSerPheAlaLeuAlaSerSerSer 673
2221 TCGGGAGGCTCAATTTGGCTTCTGATGCCCAAT.....GCTCCCAT 2264
      |||::: |||:::
674 LysSerGlnSerIleAlaMetSerGluLysGlyAsnLeuTyrAlaAspAl 690
      |||:::
2265 ACCTGATCTCTTCCACCTGGCCAAATCACCGACCTGAAGCGCGAAATTC 2314
      |||:::
690 aSerAspTyr.....IleLysSerAspLeuA 699
2315 ACGGGCGAGCTCTCAATTAATCTCACTTTGGACAGCTCCTGGGGATGATTAT 2364
      |||::: |||:::
699 snAsnGlySer.....Tyr 703
2365 GACCATGGACAGCTCACAGTATATATTCGATATAGTACAAGTATTCT 2414
      |||::: |||:::
704 GlySerAsnLysGluHisSerLeuPheIleLeuProTyrThrSerSerG 720
      |||:::
2415 TGATCTCAGACAGAGTTCAATGAATCTCTCAAGTCAATACTACTGCTC 2464
      |||::: |||:::
720 nAsnValGluLeuSerLeuAsnGluLysSerLysGlyHisThrLysGlyT 737
      |||:::
2465 TCATC.....CCAAAGGAAGCCCAACTCTGAGGAGTCTTT 2499
      |||::: |||:::
737 hrIleIleGlyTyrSerThrLeuLysAspSerGlyIleTyrGlyValTyr 753
      |||:::
2500 .....TTGTTTAAACCAAGAAA 2516
      |||:::
754 AlaGlyTyrGluAspThrLysMetGlySerThrTyrPheAspIleAsnAs 770
```

```
2517 CATTAACCTTT.....G 2527
      |||:::
770 nArgThrTyrTyrAlaGlyLeuLysTyrPheAsnThrLeuPheThrThrG 787
      |||:::
2528 AAAATGCACAGATCTTTTCATCTGCTATTACAG.....GCTGTTGAT 2568
      |||::: |||:::
787 luyGlyGlnGluValTyrIleLysAlaGlnGlyLysAlaAlaLeuIle 803
      |||:::
2569 AAGGTCGATCTGAAATCAGAAATATCCAAAC 2598
      |||::: |||:::
804 LysAsnAspLeuThrGluLysIleGlyAsn 813
      |||:::
seq_name: sp_bacteria:052781
seq_documentation_block:
ID 052781 PRELIMINARY; PRT; 1112 AA.
AC 052781;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-1998 (TrEMBLrel. 06, Last annotation update)
DE SURFACE LAYER PROTEIN.
GN SAPB2.
OS Campylobacter fetus.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=196;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CIP 5396T;
RA Guesdon J.L., Casademont I., Chevrier D.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF048699; AAC04580.1;
SQ SEQUENCE 1112 AA; 112503 MW; DA9D9C2F62602091 CRC64;

alignment_scores:
  Quality: 139.50      Length: 740
  Ratio: 0.407         Gaps: 32
  Percent Similarity: 46.351 Percent Identity: 20.270

alignment_block:
US-05-049-696-18 x 052781 ..
Align seg 1/1 to: 052781 from: 1 to: 1112

133 GTTCAATCGACCCCAATGTGCCAGAGATGAACACATCAATCAACAAT 182
    |||::: |||::: |||::: |||::: |||:::
466 ValThrValAspGlyAsnThr...LysAspLeuThrLeuSerLeuLysly 481
    |||::: |||::: |||::: |||::: |||:::
183 AAAGGACATGGTGACCCAGCATCTCTGTATCTGTTTGAAGCTACAGGA 232
    |||::: |||::: |||::: |||::: |||:::
481 sAlaGlnSerValThrGluSerSer..... 489
    |||::: |||::: |||::: |||::: |||:::
233 ACGGATTTTATTTCAAAAATGTGGCATTTTGATTCCTGAAACATGGAAG 282
    |||::: |||::: |||::: |||::: |||:::
490 .....PheLysAsnIleAlaLysThrValThrGluSer..... 500
    |||::: |||::: |||::: |||::: |||:::
283 ACAAGGCTGACTATGTGAGACCAAACTTGAGACCTACAAAAATGCTGA 332
    |||::: |||::: |||::: |||::: |||:::
501 .....AsnGlyAsnValGluThr.....ValAs 508
    |||::: |||::: |||::: |||::: |||:::
333 TGTCTGTTGTTGCTAGCTCTACTCTCCAGGT.....AATGATGAAC 373
    |||::: |||::: |||::: |||::: |||:::
508 nIleLeuAlaAsnAsnAlaThrAspLysAlaValThrIleAsnAspGlu 525
    |||::: |||::: |||::: |||::: |||:::
374 CCTAC...ACTGACGACATCGGCAACTGTGGAGAGAGGGTGAAGGATC 420
    |||::: |||::: |||::: |||::: |||:::
525 erLeuLysThrIleAsnPheSerValAspAspLysGlyAlaSerVal 541
    |||::: |||::: |||::: |||::: |||:::
421 CACCTCACTCTGATTTTCATTTGTCAGGAAAAAGTTAGCTGAATATGGACC 470
    |||::: |||::: |||::: |||::: |||:::
542 ArgAlaLysGlyLysIleValAlaAspLysAla..... 552
```

```

471 ACAAGGTAGGCATTGTCCATGAGTGGGCTCATCTACGATGGGAGTAT 520
552 .....
521 TTGACGAGTACAATAATGATGAGAAATTCCTACTTATCCAAATGGAAGATA 570
553 .....
571 CAACAGTAAGATTTTCAGCAGGATTTACTGGTACAAATGTAGTAAGAA 620
563 ThrAlaAla...AlaAspAlaValValGlnAlaAlaAsnAlaThrLysIl 578
621 GTGTCCAGGAGCGAGCTGTACACAAAAGATGCACATTCATAAAGTAA 670
578 eAspIleAsnAlaAlaLysAspThrValGlyLeuThrLeuGlyGlyVala 595
671 CAGGACTCTATGAA.....AAAGATGTGAGTTGTCTTC 705
595 laLysLeuThrAspLeuThrValAsnAsnLysGlyAlaPheAlaLeuThr 611
706 CAATCCCGCCAGCGGAGAGGCTTCTATA.....ATGTTTGC 743
612 GlyAlaAsnAlaThrAspLeuAspSerValLysAsnLeuSerValAsnTh 628
744 ACAACATGTTGATTCTATAGTTGAATTCGTACAGAGAACAAACCAACA 793
628 rGluGlyAlaPheSerIleAlaThrAlaThrSerLeuLysAsnLeuAsnA 645
794 AAGAAGCTCCAAACAGCAAAATCAAAATGCAAT..... 828
645 snLeuSerLeuAsnGlyValSerAlaAspLeuAsnSerValAsnValGly 661
829 .....CTCCGAACACATGGGAAGTATGTCGTGATTCGTAGGACTT 869
662 ThrAlaThrLeuAlaSerLeuGluAlaAsnIleAsnValSerGlyGluPh 678
870 TAAGAAAACCACTCCTATGACCAACAGCCACCAAAATCCCACTTCTCAT 919
678 eLysLeuGlyThr.....ThrThrAlaLysGlyAspValAspPheAsnI 693
920 TGCTGCAGATTGGACAAGAAATGCTGTTTGTAGTCTTGACAAATCTCGA 969
693 leGluAsnValGlyAlaLeuThrLeuGlyAlaIleThrSerSerThrGly 709
970 .....AGCATGGCACTGGTAAAC..... 987
710 AsnAlaSerValIleIleSerSerAlaThrGlyAsnValThrLeuGlyAl 726
988 .....CGCCTCAATCGACTGAATCAACGAG 1012
726 aValSerAlaThrGlnGlyAsnLeuThrLeuAsnAlaGlyAsnThrLeuG 743
1013 GCCAGCTTTTCTCG.....CTGCAGACAGTTGAGCTG 1044
743 lyAsnIleThrIleGlyAlaLeuAlaGlyAspIleValSerValAspLeu 759
1045 GGGTCCTGGGTGGATGGTGCACATTTGCACAGTCTCTGCCCATGTACAA... 1092
760 GlyGlyValLeuGlyThrIleAsnSerAlaSerGlyAsnLysValGluIl 776
1093 .....AGTGAACATCATACAGATAAAGTGGCAGTGCACAGGCACAC 1135
776 eThrSerAsnGluValThrThrValGlySerGluIleSerLysAsnValV 793
1136 TCCGCAAAAGATTACCTGCAGCAGCTTCAGGAGGAGCTCCATCTGCAGC 1185
793 alGlu.....IleThrAlaAlaAlaGlyGlyThrAspLeuAsnAla 806
1186 GGGCTTCGATCGGCAATTTACTGTGATAGGAGAAATATCCAACTGATGG 1235
807 .....GlnValIleGlyGlyAlaAlaAlaAspAspAl 817
1236 ATCTGAAATTGCTGCTGACGGATGGGAGAGACACATATAGTGGGT 1285

```

```

aLeuThrIleIleGlyLysGlyAspThrGlnThrIleThrAlaSerGly. 833
1286 GCTTTAAGAGGTCAAAACAAAGTGGTCCATCATCCACACAGTGCCTTTG 1335
836 .....AspLeuSerGlyGlyThrLeu...ThrLeuThrLeu 844
1336 GGGCCCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAAATGACAGGAGG 1385
845 ThrAspAlaThrLysLeuSerSerLeuAspIleSerGlyValLysGlyLe 861
1386 TTTACAGACATATGCTTCAGAT.....CAAGTTCAGAACAAATG 1423
863 userAlaAlaThrAlaIleAspLeuLysAsnValSerValGluAsnLysL 878
1424 GCCTCATTTGATGCTTTTGGGCCCTTTTCATCAGGAAATGGAGCTGTCTCT 1473
878 euIleValAspIleGlnGly.....SerAspAlaAlaGluThrIleThr 892
1473 CAGGCTCCATCCAGCTTGAGAGTAAAGGATTACCTCCACAGACGCA 1523
893 AlaAsnSerThrSerAlaThrLeuThrAlaIleThrLeuSerGlyAspLe 909
1524 GTGGATGAATGGCACAGTGATCGTGGACAGCACCGTGGGAAAGGACACTT 1573
909 uGlyGlyGlyAlaAsnThrVal.....ThrValAlaProAspAlaA 923
1574 TGTTCCTTATCACCTGGACAAACGCGCTCCCAAAATCCTTCTCTGGGAT 1623
923 laAlaValAlaIleThrIleAspLeuSerGlyLeu.....Ser 936
1624 CCCAGTGGACAGAGCAAGGTGGCTTTGTAGTGACAAAACACCAAAAT 1673
937 AlaThrGlyGlyThrLeuSerGlyThrIleThrHisAsnAlaAlaGlnTh 953
1674 GGCTACTCTCCAAATCCACAGCAATTTGCTAAGTGTGGCACTTGGAAATACA 1723
953 rAlaLeuThrThrIleLysGly..... 960
1724 GTCTGCAAGCAAGCTCACAAACCTTGACC..... 1752
961 .....SerAlaGlyAsnAspThrIleThrIleGlyLysValAsnAspGly 975
1753 CTGACTGTCAGCTCCGCTGGCTCCAAAT..... 1779
976 LeuThrValThrGlyGlyAlaGlyAsnAspValPheAsnValThrAlaAl 992
1780 .....GCTACCTGCTCCCA...ATTACAGTGAATCTCCAAA 1813
992 aLysIleValThrAlaAspThrProGluHisAlaThrIleThrAspPheS 1009
1814 CGAACAGGACACCAAGCAAAATTCGCCAGCCCTCTGGTAGTTTATGCAAT 1863
1009 erAlaGlyAspSerIleLysPheAlaAlaSerValThrAlaThrGlyAsn 1025
1864 ATTGCCAAGGAGCTCCCAATTTCTCAGGCGCAGTGTGCAGAGCCCTGAT 1913
1026 ValGlyThrValAlaGlyAspThrLeuLysAlaAlaIleLysAlaAlaIl 1042
1914 T.....GAAT 1918
1042 eAlaLeuThrAspLysAlaProGlyIleThrSerAlaAspLysGluThrT 1059
1919 CAGTCAATGGAAACACAGTTACCTTGGAA.....CTACTGGATAAT 1959
1059 hrValThrGlyPheThrTyrAsnGlyAspAsnTyrLeuPheThrAsnAsn 1075
1960 GGACAGGTGCTGATGCTACTAAG...GATGACGGTGTCTACTCAAGTA 2006
1076 AlaAsnGlySerAspSerThrThrValAspValLeuValLysLeuTh 1092
2007 TTTACACACTTATGACACGAATGGTAGATACAGTGTAAAGTGGGGCTC 2056

```



```
1007 AAGCAGGCCAGCTTTCTCTGCTGACAGACAGTTGAGCTGGGGTCTCTGGGTT 1056
      |||:|||||:|||||
394 .....ThrAlaGlnLeuSerSerSer 401
1057 GGGATGGTGACATTGACAGTGCTGCCCATGTCACAAAGTGAACTCATACA 1106
      |||:|||||
402 LysThrLeuThr.....SerThrGln 408
1107 GATAACAGTGGCAGTGACAGGACACACATCGCCAAAGATTACCTGCAG 1156
      |||:|||||
408 nThrThrGlnGlySerPro...SerThrValSerGlnThrThrProGlyV 424
1157 CAGCTTCAGGAGGACGTCCATCTGACCGCGGCTTCGATCGGCATTTACT 1206
      |||:|||||
424 alSerSerAlaSerThrGlyMetThrThrSerGlnAlaSerLeuArgSer 440
1207 GTGATTAGGAAGAAATATCCAACTGATGATCTGAAATGTGCTGCTGAC 1256
      |||:|||||
441 ThrGlnAsnSerValThrSerThrProGlySerPheThrValThrSerTh 457
1257 GGATGGGAAGACACACTATAAGTGGGTGCTTTAAGCAGGTCAACAAA 1306
      |||:|||||
457 rProThrIleThrSerThr.....G 464
1307 GTGTGCCATCATCCACACAGCTGCTTGGGGCCTCTGCAGCTCAAGAA 1356
      |||:|||||
464 InGlyAlaMetAlaSerSerSerAsnSerProThrSerThrGlnAla 480
1357 CTAGAGGAGCTGTCACAAATCAGAGGAGTTTACAGACATATGCTTCAGA 1406
      |||:|||||
481 IleSerSerSerThrValThrThrValAlaSerSerThrIleAlaProTh 497
1407 TCAAAGTTCGAACAAATGCCCTCATGTATGCTTTTGGGGCCCTTTCATCAG 1456
      |||:|||||
497 rSerSerGlnGluSerSerSerIleAlaSerSerSerAlaProSerGlns 514
1457 GAATGGAGCTGCTCTCAGCGCTCCATCCAGCTTGACAGTAAGGGATTA 1506
      |||:|||||
514 erThrThrValIleThrValSerThrThrAlaThrValSerSerGly... 529
1507 ACCCTCCAGAACAGCAGCTGATGAATGGCAGTGCATCGTGGACGAC 1556
      |||:|||||
530 .....GlnSerThrSerThrPheGlyThrThrIleGlyGlnGlySerSe 544
1557 CGTGGAAAGACACTTTGTTCTTATACCTGGACACAGCAGCTCCGCC 1606
      |||:|||||
544 rPheGlySerSerThrIle.....GlnThrThrGlnGlyThrS 557
1607 AAATCCCTCTCTGGGATCCAGT..... 1629
      |||:|||||
557 erSerPheAlaProIleProSerThrAlaGlySerSerSerGlnThrPro 573
1630 .....GGACAGAACAGAGTGGGCTT 1649
      |||:|||||
574 GlySerMetSerSerThrGlyThrThrValGlyGlnMetSerSerSerPh 590
1650 TGTAGTGACAAAACACAAAATG...GCCTACTCTCAAAATCCACAGCA 1696
      |||:|||||
590 eGlnProThrAlaProThrSerLeuGlyThrIleMetThrThrProGlyT 607
1697 TTGCTAAGTTGGCACTTGGAAATACACTCTGCAAGCAAGCTCAACACC 1746
      |||:|||||
607 hrSerSerIleProSerIleSerThrSerValAsnSerGlySerSerThr 623
1747 TTGACCCCTGACTGTACATCCCGTGGCTGCTCAATGCTACC..... 1785
      |||:|||||
624 IleGlySerThrValThrGlnAlaProSerSerSerSerThrSerMetGlyPr 640
1786 .....CTGCCTCCAA 1795
      |||:|||||
640 oSerProSerGlnSerThrAlaGlySerThrMetThrSerAlaProProv 657
```

```
1726 TTACAGTGACTTCCAAACGAACAAG...GACACCAGCAAAATTCCTCCAGC 1842
      |||:|||||
657 alThrSerSerSerAlaAsnThrGlySerThrSerSerGlyThrThr 673
1843 CCTCTGTAGTTTATGCAAAATATTCGCAAGGAGCCTCCCAATTCCTCAG 1892
      |||:|||||
674 ValSerValGlnThrThrGlnValSerSerThrThrSerProValAlaSer 690
1893 GGCCAGT.....GTCACAGCCCTGATTGAATCAGTGAATGAAGAAAACAG 1936
      |||:|||||
690 rSerSerSerGlnMetThrSer...ThrGlnGlnProSerGlySerSers 706
1937 TTACTCTGGAACACTACTGGATAATGGACAGCAGTGTGCTGCTACTAAGAT 1986
      |||:|||||
706 erSerIleGlySerThrValAsnGlnGlySerSerValThrGln 722
1987 GACGGTGTCTACTCAAGGTATTTACACAACTTATGACACGAATGGTAGATA 2036
      |||:|||||
723 ProProAlaSerSerArgSerThrAlaSerGlnGlySerSerSerAlaGl 739
2037 CAGTGTAAGTGGCGGCTCTCTGGAGGAGTTAACGCA..... 2073
      |||:|||||
739 nProIleAlaSerSerSerThrMetGlySerThrAlaGlySerSerSerp 756
2074 .....GCCAGCAGGAGTGTATACCCACAGCAGTGGAGCA 2109
      |||:|||||
756 roGlnProThrAlaSerSerThrSerValProSerSerThrGlyAla 771
seq_name: sp_vertibrate:Q9W6V5
seq_documentation_block:
ID Q9W6V5 PRELIMINARY; PRT; 1406 AA.
AC Q9W6V5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE SUPPORTING-CELL ANTIGEN PRECURSOR (EC 3.1.3.48).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCHI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RA Kruger R.P., Goodyear R.J., Legan P.K., Marchol M., Raphael Y.,
EA Cotanche D.A., Richardson G.P.;
RT "The supporting-cell antigen: a receptor-like protein tyrosine
phosphatase expressed in the sensory epithelia of the inner ear.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ238216; CAB41885.1; -
DR HSSP; P18052; LYFO.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; Tyr_prot-phptase.
DR Pfam; PF00041; fn3; 9.
DR Pfam; PF00102; i_phosphatase; 1.
DR PRINTS; PR00700; PRTVPHPTASE.
DR SMART; SM00060; FN3; 7.
DR SMART; SM00194; PTPc; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase; Signal.
FT SIGNAL 1 24 POTENTIAL.
SQ SEQUENCE 1406 AA; 154196 MW; 416387A29B3EC7C2 CRC64;
```

alignment_scores:

```
Quality: 139.00 Length: 556
Ratio: 0.599 Gaps: 21
Percent Similarity: 41.727 Percent Identity: 19.604
```

alignment block:

US-09-049-696-18 x Q9W6V5

Align seg l/1 to: Q9W6V5 from: 1 to: 1406

1366 CTGTCACAAAATCACAGAGGGTTTACAGACATATGCTTCAGATCAAGATTCA 1415
|||||:::|||||::: |||| :::::||:::|||||:::|||||:::
34 LeuLysAsnValThrGlucIumMetGlyThrSerSerAsnAspGluLeuSe 50

1416 GAACAATAGCCCTCATGTGATGCTTTTGGGGCCCTTCATCAGGAATGGAG 1465
: ||||::: |||||::: |||||::: |||||::: |||||:::
50 rValAsnAla.....ThrSerGlyAsnArgA 59

1466 CTGCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAACCCTCCAG 1515
:::|||||::: :::::|||||::: |||||::: |||||::: |||||:::
59 rgLeuSerGlu...AspValSerLeuProGlyArgAlaMetSerAspGln 74

1516 AACAGCCAGTGGATGAATGCACAGCTGATCGTGGACAGCACC.....GT 1559
|||||::: |||||::: |||||::: |||||::: |||||:::
75 AsnSerValAlaGlnProArgAlaValLeuAspLeuLysThrGluTrpVa 91

1560 GGGAAAGGACACTTTGCTTTCATCACCTGGACACGCCCTCCCCCAA 1609
|||||::: |||||::: |||||::: |||||::: |||||:::
91 lglyValThrSerValAsnLeu...ThrTrpThrValAsn..... 103

1610 TCCTTCTCTGGATCCAGTGGACAG..... 1635
|||
104AspThrAlaSerAspSerTrpThrTyrrgileGluVal 116

1636 AAGCAAGTGGCTTTGTAGTGGACAAAAACACCAAAATGCCCTACCTCCA 1685
|||||::: |||||::: |||||::: |||||::: |||||:::
117 ArgAsnGlySerSerIleAsnAsnLysThrSerAspIleThrAspAlaGl 133

1686 AATCCCAGGCAATTGCTAAGTTGGCACTTGGAAATACAGTCTCCAAGCA. 1734
: ||| |||||::: |||||::: |||||::: |||||::: |||||:::
133 uilleThrGlyLeuIleProGlyThrLeuTrpThrPheThrValPheAlav 150

1734 1734

150 alAlaAlaAspGlyGlnThrAlaGlyGluGlyAlaSerIleSerLeuTrp 166

1735AG 1736

167 ThrLysProSerProValLeuAspLeuLysAlaGluTrpValGlyValTh 183

1737 CTCACAAACCTTGACCCCTGACTCTCAGTCCCGTGGTCCAAATGCTACCC 1786
: ||| :::::|||||::: |||||::: |||||::: |||||:::
183 rSerValAsnLeuThrTrpThrValAsnAspThrAlaSerAlaSerTrp 200

1787 TGCCTCCAAATACAGTACTTCCAAACAGNACANGACACCACCAATAATC 1836
||| |||||::: |||||::: |||||::: |||||:::
200 hrTyrrgileGluValThrSerAspSerSerIleAspSerLeu..... 214

1837 CCAGCGCTCTGTAGTTTATGCAAAATATTCGCAAGGAGCCCTCCCAAT 1886
: |||||::: |||||::: |||||::: |||||::: |||||:::
215ThrSerSerVa 218

1887 TCTCAGGGCCAGGTGCACAGCCCTGATTGAATCAGTGAATGGAAAAACAG 1936
: |||||::: |||||::: |||||::: |||||::: |||||:::
218 lThrMetAlaGluIleThrGlyLeuIlePro.....GlyThrLeu 232

1937 TTACCTTGGNACTACTGGATAATGGAGCAGGTGCTGATGCTACTAAGGAT 1986
: |||||::: |||||::: |||||::: |||||::: |||||:::
232 yrSerPheLysvalPheAlaValAlaAlaAsp...AsnArgThrGluAla 247

1987 GACGGTGTCTACTCAAGGATATTCACA..... 2013
|||||::: |||||::: |||||::: |||||::: |||||:::
248 AspGlyAlaSerIleSerLeuTrpThrLysProSerProValLeuAspLe 264

2014 2014

264 uLysAlaGluTrpValcIyValThrSerValAsnLeuThrTrpThrVala 281
|||||::: |||||::: |||||::: |||||::: |||||:::

547 TTCTACTTATCAATGGGAAGAAATAACAACGATAGAATGTTCAGCAGGTAT 596
||||| :||::|| | || :: |:: ||::|
122 PheTyAsnGluHisSerGlyIleLeuSerPheMetalArGserGlyVa 138

597 TACTCGT ACAAATGTAGTAAGAAGTGTCAGGAGCCA 634
|| | |:|::|::|::|::|::|::|::|::|:
138 iGluGlySerLeuThrLeuSerAsnIleLysMetThrGlyAspGlyGlyA 155

635 GCTGTATCACAAA... AGATGCCCATTTCAAATAAGTAACAGGACTCTAT 681
||::|::|::|::|::|::|::|::|::|::|::|:
155 laileTySerGlnGlyGluLeuLeuPheThrAspLeuThrGlyLeuThr 171

682 GAANAAGATGTGATTGTTGTTCCATCCCGCCAGGCAGGAGAGGCTTC 731
:||:||||: ||::|::|::|::|::|::|::|:
172 IleGInGly..... AsnLeuSerGlnLeuSerGlyGlyGI 183

732 TAATAATGTTGCACAACATGTTGATTCTATAGTTGAATCTGTACAGAAC 781
|||::|::|::|::|::|::|::|::|::|::|:
183 yllePheGlyGly..... SerThrIleserPhe..... 192

782 AAACCACAACAAGAAGCTCCAACAAGCAAAAAATCAAAAATGCAATCTC 831
::: |||||
193 SerGlyIleAsnGlnAlaThrPheSer 201

832 CGAACCATCGGAAGTAGTCGGTGATTCTGAGNCTTTAAGAAACCAC 881
::::||| ||||::|::|::|::|::|::|::|:
202 SerAsnThrAlaGluValProGlu..... GluThrTh 213

882 TCCT, ATGACACACAGCAGCACCAATCCCA 910
|||::|::|::|::|::|::|::|::|::|::|:
213 rProAsnProAsnProGlyThrGlnThrThrSerGlnProSerProt 230

911 CCTTCTCATGTCGAGATTGGCAAGAATVGTGTTTTAGTCCTTWGAC 960
|| ::::::::::: |||
230 hr..... SerLysValGlnSerLeuPheThrTy 239

961 AATCTGGAAGCATGGCAGCTGGTAACCGCTCAATCGACTGAATCAAGC 1010
:::||||::|::|::|::|::|::|::|::|:
240 SerSerSerThrGlnAlaAsnGlyasn..... 248

1011 AGGCCAGCTTTTCTCTGCTGCAGACAGTTGAGCTGGGTCTCTGGGTGGGA 1060
248 248

1061 TGGTGACATTTACACTGCTCCCCATGTACAAAGTGAATCATACAGATA 1110

248 248

1111 AACAGTGGCAGTGACAGGACACACTGCCAAAAGATTACCTGCAGCAGC 1160
|||::|::|::|::|::|::|::|::|::|::|:
249 GlyAlaAspSerGlnThrProSerHisiLys..... ProGL 260

1161 TTCAGGAGGAGCGTCCATCTCGACGGGCTTCGATCGGCATTTACTGTGA 1210
|||::|::|::|::|::|::|::|::|::|::|:
260 yseryGlyGlyAlaIleTyAlaThrGly..... AspLeuthrIle. 273

1211 TTAGCAAGAAMATCCAACACTGATGNAUCTGAAATVTGCTGCTGACGGAT 1260
:::||||::|::|::|::|::|::|::|::|:
274 SerASPserGlnGlullelvalPhe..... 281

1261 GGGGAAGACAACACTATAAGTGGGTGCTTTTAAACAGAGTCAACAACATGG 1310

282 SerValasnlunylsAlaserLysaspGL 290

1311 TGCCCATCATCCACACAGTCGCTTTGGGGCCCCTCTGCGAGCTCAAGAAGTAG 1360

290 yGlyAlaIle..... 293

1361 AGGAGCTGTCCAAAATGACAGAGGTTTACAGACATATGCTTCATGACAA 1410

294PheAlaGluLysAsnValSerPheGluAsnIleThrThrLeuLys 308

1411 GTTCAGAACAAATGGCCTCATTTGATGCTTTTGGGCGCCTTCATCAGGAAA 1460

309 ValGlnAsnAsn.....GlyAlaGluGluLysGlyGly 319

1461 TGGAGCTGCTCTCAGCGCTCCATCCAGCTTGACAGTAGAGGATATACCC 1510

319 yGlyIleThrAlaSerGlyAspLeuSerIleGlnSerSerLysGlnSerL 336

1511 TCCAGAACAGCAGTGGATGAATGGCAGAGTGCATGCTGGACAGCACCGTG 1560

336 euPheAsnSer..... 339

1561 GGAAGGACACTTGTCTTATCACCCTGGACACAGCAGCGCTCCCCAAAT 1610

339 339

1611 CCTTCTCTGGATCCAGTGGACAGCAAGGTGGCTTTGTAGTGACA 1660

340AsnThrSerLysGlnGlyGlyAlaLeuThrIleGlu 353

1661 AAACACCAAAATGGCCTACCTCCAAATCCCGACGATGCTTAAGTTGGC 1710

353 LyAsnValAspPheLysAspLeuGluIleArgIle..... 365

1711 ACTTGGAAATACAGTCTCCAGCAAGCTCACAACCTTGACCCGTACTGT 1760

366LysTyAsnLysSerGlyThrPheGluThrLysLysValThrLe 380

1761 CAGTCCCGTGGTCAATGCTACCTGCCCTCCAAATACAGTACAGTCTCCA 1810

380 uSerLeuProGluAlaGlnThrAsnLysSerValThrAlaAlaSerG 397

1811 AAACGAACAAGCACACCAAGAAATCCCGACCCCTCTGGTA..... 1851

397 InSerGlyProAsnThrThrProThrProThrProValThrAlaLys 413

1852GTTATGCAAAATATTCGCAAGAGCGCTCCCAATCTCTCAG 1892

414 GlyGlyGlyLeuThrThrGluLysAsnLeuSerIleSer..... 426

1893 GCCAGTGTACAGCCCTGATGATCAATCAGTCAATCGAAGAAACAGTTACCT 1942

427AsnIleThrGlyIleIleGluIleThrAsnAsnLysAlaThr.... 440

1943 TGGAACTACTGGATAATGGACAGAGTGCT..... 1971

441AspValGlyGlyAlaTyValLysGlyThrLeuThr 453

1972GATGCTACTAA 1982

454 CysLysAspSerHisArgLeuGlnPheGlnLysAsnSerSerGluLysLy 470

1983 GGATCAGCGTGTCTACTACAGTATTTACAACTTATTCACACGAAATCGTA 2032

470 sGlyGlyGlyLeuThr.....ThrGluAspThr..... 479

2033 GATACAGTGTAAAGTGGCGGCTCTGGAGGAGGTAAACGACCCAGACGG 2082

480IleThrLeuSerAsnLeuThrGlyLysThrLeuPheGlnGlu 493

2083 AGAGTGATACCCACAGCAGTGGACACTGTACATACCTGGCTGGATTGA 2132

494 AsnThrAlaLysGluGluGlyGlyLeuTyIleGlnGly..... 507

2133 GAATGATGAATACAAATGCCAACAGACCTGGAATTAATAAGGATG 2182

508AspA 509

	seq_documentation_block:	
ID	YPR4_CAEEL	STANDARD;
AC	Q20059;	PRT; 686 AA.
DT	15-JUL-1998	(Rel. 36, Created)
DT	15-JUL-1998	(Rel. 36, Last sequence update)
DT	15-JUL-1998	(Rel. 36, Last annotation update)
DE	HYPOTHETICAL 77.0 KDA TRP-ASP REPEATS CONTAINING PROTEIN F35G12.4 IN CHROMOSOME III.	
DE	F35G12.4.	
GN	Caenorhabditis elegans.	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;	
OC	Rhabditidae; Peloderinae; Caenorhabditis.	
OX	NCBI_TaxID=6239;	
RN	{1}	
RN	{1}	

271 pSerSerPheThrThrValTyrSerAlaGlyLysAspLysMetValValL 288
1583 TCACC.....TGACAAACGAGCCTCCCAAAATCCTTCTCTGGGATCCC 1626
288 yThrProLeuTyrAspPheThrLysSerGlnLeuLeuPheLysGluGlu 304
1627 AGTGGACAGAGCAAGGTGGCTTTGTAGTGGACAAACACCAAAATATGGC 1676
305 AlaProValLysLys.....LeuLeuLeuSe 313
1677 CTACCTCCAAATCCAGGACATTCCTAAGTTGGCATTGGAA..... 1719
313 rGluLysAspAsnProValSerLeuTyrValGlyThrTyrLysSerAspI 330
1720TACAGTCTCCAGCAAGCTCACAAACCTTGACCCGCTGCTC 1761
330 leLysArgTyrPheSerIleArgProSerAlaGln.....LeuSerIle 343
1762 ACGTCCGCTGGTCCCAATGCTACCTCCCTCCAATTACAGTGACTCCAA 1811
344 GlyLysAspGluAspGlyProSer.....ThrSerAs 354
1812 AACGAACAGGACACCAACAAATTCCTCCAGCCCTCTGGTAGTTATGCAA 1861
354 nAlaAsnHisSerValSerAlaSerSerProProValThrPheLysT 371
1862 ATATTCCCAAGGAGCCTCCCAATTCCTCAGGGCCAGTGTACAGCCCTG 1911
371 yrIleArg..... 373
1912 ATTGAATCAGTGAATGAAACACAGTTTACCTTGGCACTACTGATATGG 1961
374 ValLysAspGlnLysGlyGlnSerThrProGluLeuValIle..... 388
1962 AGCAGGTGCTGATCTACTAAGTAGCAGGTGTCTACTCA.....AGGT 2005
389 ProGlyAlaProAlaIleLysLysHisAlaMetLeuSerAspLysArgH 405
2006 ATTTCACAATTATGACACCAAGTGGTAGATACAGTGTAAAGTCGGGCT 2055
405 isValLeuThrArgAspSerAspGlyVal.....Ala 416
2056 CTGGGAGGAGTTAAGCAGCAGCAGCAGG..... 2082
417 LeuTyrAspValLeuAlaIleArgLysIleLysAspTyrGlyLysArgII 433
2083AGAGTGATACCCAGCAGGTGGAGGAGCTGTACATACCTGGCT 2125
433 ePheGluGluValValAspGluAsnSerArgGlnValTyrIleProSerT 450
2126 GG.....ATTGAGATGATGAA 2142
450 rPheValValAspSerLysSerGlyMetLeuGlnIleThrLeuAspGlu 466
2143 ATA.....CAATGGAATCCACCAAGACCTGAA...ATTAATAA 2177
467 LeuAspAlaLeuSerSerTyrPheSerLysAspAlaGlyPheAspAs 483
2178 GGATGATGTTTCAACACAAAGTGTGTTTTCAGCAGAACATCTCCGGAG 2227
483 pAsnAspArgGluThrLysGlnIleIle.....ValAsnTyrGlyG 497
2228 GCTCATTTGTGGCTTCTGTATGTCCTCCCAATGCTCCCATACCTGATCTTTC 2277
497 lyMetMetLeuArgSer.....LeuPheGluArgTyr 507
2278 CCACCTGGCCAAATCACCGACCTGAGGGGAAATTCACGGGGGCGAGTCT 2327
508 ProProCysLysMetThrAsnValAspAla..... 517
2328 CATTAATCTGACTTGGACAGCTCTCTGGGATGATATGACCATGGGAACAG 2377
518AlaAspAlaAspValGlnLysAlaThr. 527

2378 CTCACAAGTATATCATTCGAATAAGTACAAAGTATCTTGTATCTCAGAGAC 2427
528LeuAsnPheIleSerLeuProGlu 535
2428 AAGTTCATGAATCTCTCAAGTGAATACTACTCTCTCATCCCAAGGA 2477
536 His.....ThrProLeuIleIleCysGI 543
2478 AGCCAACCTCTGAGGAAGTCTTT.....TTGTTTAAACCAAGAAACATTA 2521
543 uGlyAsnGlyArgProLeuTyrArgLeuLeuValGlyAspAlaGlyLysG 560
2522 CTTTGTGA...AATGGCACAGATCTTTTCATTCCTATTACAGCTGTGTGAT 2568
560 lupheGluAlaAsnGluLeuAlaGlnIleAlaProMetTyrPvalIleAsp 576
2569 AAGTGTGATCTGAAATCAGAAATATCCAAACATTCACGAGTATCTTTGTT 2618
577 AlaIleGlu...ArgAsnGlnLeuProLysPheAsnLysMetProPheTyr 592
2619 TATT.....CCTCCACAGACTCCGCCAGACACACCTAGTCTCTGATGAAA 2662
592 rLeuLeuProHisProSerThrAsnProLysGlnProLysLysAspArgL 609
2663 CGTCTGCT 2670
609 euSerAla 611
seq_name: SwissProt_39:IGA4_HAEIN
seq_documentation_block:
ID IGA4_HAEIN STANDARD; PRT; 1849 AA.
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE IMMUNOGLOBULIN A1 PROTEASE PRECURSOR (EC 3.4.21.72) (IGAL PROTEASE).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:NHTI HK61;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Killian M.;
RT "A comparative genetic study of serologically distinct Haemophilus influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC - FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC - PRODUCING INTACT FC AND FAB FRAGMENTS.
CC - CATALYTIC ACTIVITY: CLEAVAGE OF IMMUNOGLOBULIN A MOLECULES AT
CC - CERTAIN PRO-1-XAA BONDS IN THE HINGE REGION. NO SMALL MOLECULE
CC - SUBSTRATES ARE KNOWN.
CC - SUBCELLULAR LOCATION: SECRETED.
CC - DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC - SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC - OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC - DOMAIN: THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC - SIMILARITY).
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M87491; AAA24968.1; -

DR MEROPS: S06.001; -
 DR InterPro: IPR000710; Iga_S6.
 DR Pfam: PF02395; IGA1; 1.
 DR PRINTS: PR00921; ICASERPTASE.
 KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.
 FT PROPEP 1022 1849 HELPER PEPTIDE (POTENTIAL).
 FT ACT_SITE 299 299 PROBABLE.
 SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;

alignment_scores:

Quality: 136.00 Length: 956
 Ratio: 0.346 Gaps: 53
 Percent Similarity: 41.109 Percent Identity: 21.339

alignment_block:

US-09-049-696-18 x IGA4_HABIN ..

Align seg 1/1 to: IGA4_HABIN from: 1 to: 1849

```

349 TCTACTCTCCAGGT.....AATGATGAACCTACACTGACGAG..... 387
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
279 SerAspProLysGlyLeuSerGlnAspProLeuThrAsnTyrAlaVala 295
388 .ATGGGCAACTGTGA.....GAGAAGGGTG 412
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 lLeuGlyAspSerGlySerProLeuPheValTyrAspArgGluLysGlyL 312
413 AAAGNATCCACCTC...ACTCTGTATTCATTGCGAGGAAAAAGTTAGCT 459
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 ysrTrpLeuPheLeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLys 328
460 GAATATGGACCAAGGTAGGACATTTCCTCATGAGTGGGCTCATCTAGC 509
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
329 SerTrp...GlnGluTrpAsnIleTyrLysHisGluPheAla..... 341
510 ATGGGGAGTATTGACGAGTACATAATGATGAGAAATCTACTTATCCA 559
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342 .....GluLysIleTyr..... 345
560 ATGGAAGAAATCAACAGCAGTGTTCAGCAGGT...ATTACTGTGACA 606
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
346 .....GlnGlnTyrSerAlaGlySerLeuThrGlySer 356
607 AATGTAGTAAGAAGTGTGAGGAGGACGCTGTACACCAAAAGATGCAC 656
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
357 Asn.....ThrGlnTyrThrTrpGlnAlaLath 365
657 A.....TTCATAAGTAACAGGACTCTATGAAAAGGATGTGAGTTG 700
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
365 rGlySerThrSerThrIleThrGly..... 373
701 TTCCTCAATCCCGCCACGACGGAGAGGCTTCTATATGTTTGCACAACAT 750
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
374 .....GlyGlyGluProLeuSerValAspLeuThrAspGly 385
751 GTTGATTTCTATGTTGAATCTGTACAGACAAACCAACCAAGAAAGC 800
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
386 LysAsp.....LysProAsnHisGlyLysSerI 395
801 TCCAAACAGCAAAATCAAAATGCATCTCCGA..... 834
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
395 eThrLeuLysGlySerGlyThrLeuThrLeuAsnAsnHisIleAspGlnG 412
834 ..... 834
412 lYlaGlyGlyLeuPhePheGluGlyAspTyrGluValLysGlyThrSer 428
835 .....AGCACATGGGAAGTATCCGTTGATCTGAGGACTTTAAGAAAC 878
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
429 AspSerThrThrTrpLysGlyAlaGlyValSerValAlaAspGlyLysTh 445

```

```

879 CACTCTCTATGACAACACAGCCACCAATCCACCTTCTCA...TTGTGCG 925
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
445 r.....ValThrTrpLysValHisAsnProLysTyrAspArgLeuAlaL 460
926 AGATTGCACAAGAATTGTGTGTTAGTCTCTTGAC.....AAA 963
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
460 ysrIleGlyLysGly.....ThrLeuValValGluGlyLysGlyLysAsn 474
964 TCTGAAAGCATGGGACTGTGTAACCGCTCAATCGACTGAATCAAGCAGG 1013
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
475 GluGlyLeuLeuLysValGlyAsp.....G 483
1014 CCAGCTTTTCTCTGCTGCAG.....ACAGTTGAGCTGG 1045
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
483 yThrValIleLeuLysGlnLysAlaAspAlaAsnLysValGlnAlap 500
1046 GCTCTGGGTTGGGATGGTGACATTTGCACAGTCTGCTGCCCATGTACAAAGT 1095
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
500 heSerGlnValGlyIleValSerGlyArgSerThr..... 511
1096 GAACCTACATACAGATAACAGTGGCAGTGACAGGACACACATCGCCAAAAG 1145
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
512 .....LeuValLeuAsn.....AspAspLysGlnValAspProAsnSe 524
1146 ATTACCTGCGACAGCTTCAGAGGAGCCTCCATCTGCGAGCGGCTTCGAT 1195
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
524 rIleTyrPheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnSerL 541
1196 CGGATTTACTGTGATTAGGAAGAAATATCCAACCTGATGGAATGAAAT 1245
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
541 euThrPheAspHisIleArgAsn.....IleAspAspGlyAlaArgVal 555
1245 GTG.....CTGCTGACGATGGGAGACACACTATAAGTGGG..... 1284
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
556 ValAsnHisAsnMetThrAsnThrSerAsnIleThrIleThrGlyGluse 572
1284 ..... 1284
572 rLeuIleThrAsnProAsnThrIleThrSerTyrAsnIleGluAlaGlnA 589
1285 .....TGC 1287
589 spAspAspHisProLeuArgIleArgSerIleProTyrArgGlnLeuTyr 605
1289 TTTAAGAGGTCAACAAGTGGTCCCATCATCCACACAGTCGCTTTGGG 1337
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
606 PheAsnGlnAspAsnArgSer.....TyrTyrThrLeuLysLysG 619
1338 GCCCTCTGCGCTCAAGAACTAGAGGAGCTGTCCAAAATGACA..... 1380
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
619 yAlaSerThrArgSerGluLeuProGlnAsnSerGlyGluSerAsnGluA 636
1381 .....GGAGGTTTACAGACATATGCTTCAGATCAAGTTCCAG 1416
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
636 snTrpLeuTyrMetGlyArgThrSerAspGluAlaLysArgAsnValMet 652
1417 AAC.....AATGGCTCATTTGATGCTTTTGGGGC 1445
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
653 AsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrPheGly.. 668
1446 CCTTTTCATCAGAAATGGAGCTGTCTCTCAGCGTCCATCCAGCTTGAGA 1495
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
668 .....GluGluGluThrLysAlaThrGlnAsnGlyLysLeuAsnValThrP 684
1496 GTAAGGGATTAAACCTCCAGACACAGCCAGTGGATGAATGGCAGCTGATC 1545
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
684 heAsnGlyLysSerAspGlnAsnArgPheLeuLeuThrGlyGlyThrAsn 700
1546 GTG.....GACAGCACCGTGGGAAAGACACTTTGTTTCTTATCACCTG 1589
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
701 LeuAsnGlyAspLeuAsnValGluLysGlyThrLeuPheLeuSerGlyAr 717

```


4-SULFATE BRIDGE TO THE THEIR C-TERMINAL ASPARTATE (BY SIMILARITY).
 -|- SIMILARITY: BELONGS TO THE ITH FAMILY.
 -|- SIMILARITY: CONTAINS 1 VWFA DOMAIN.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed, usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: D89286; BAA13939.1; ..
 InterPro: IPR002035; VWFA.
 Pfam: PF00092; vwa; 1.
 SMART: SM00327; VWA; 1.
 PROSITE: PS0234; VWFA; 1.
 KW Serine protease inhibitor; Repeat; Signal; Multigene family; Glycoprotein.

FT	SIGNAL	1	18	POTENTIAL.
FT	PROPEP	19	54	BY SIMILARITY.
FT	CHAIN	55	702	INTER-ALPHA-TRYPsin INHIBITOR HEAVY CHAIN
FT	PROPEP	703	946	H2; BY SIMILARITY.
FT	DOMAIN	308	468	VWFA.
FT	CARBOHYD	118	118	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	263	263	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	445	445	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	578	578	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	BINDING	702	702	CHONDROITIN 4-SULFATE, CROSS-LINK SITE (BY SIMILARITY).

SQ SEQUENCE 946 AA; 106580 MW; CA8BF565458E7B2E CRC64;

alignment_scores:
 Quality: 135.50 Length: 1008
 Ratio: 0.320 Gaps: 54
 Percent Similarity: 42.063 Percent Identity: 20.139

alignment_block:

US-09-049-696-18 x ITH2_MESAU ..
 Align seg 1/1 to: ITH2_MESAU from: 1 to: 946

82	CTGAGTAATCTACTATCAG.....CTGACAAACAATGGGTATGAAG	125
85	MetAlaThrThrIleGlnSerLysLeuValAsnAsnSerProGlnSe	101
126	CATTGCTGTCGCAATCGACCCCAATGTCGAGAGATGAA.....	165
101	rGlnAsnValAlaPheAspValGlnIleProLysGlyAlaPheIleSerA	118
166ACACTCATCAACAATAAAGGAC	189
118	snPheThrMetThrValAsnGlyIleThrPheThrSerThrIleArgGlu	134
190	ATGGTGACCCAGCATCTGTGAT...CTGTTGAAGCTACAGGAAGCG	236
135	LysThrValGlyArgAlaLeuTySerGlnAlaArgAlaLysGlyLys..	150
237	ATTTTATTTCAAAATGTTGCCATTTTGTATCTCGAAACATGGAACAA	286
151T	151
287	AGGCTGACTATGTGAGACCAAACTTGAGACCTACAAAATGCTGATGT	336
151	hrAlaGlyTrpValArgSerArgThrLeuAspMetGluAsnPheAsnThr	167
337	CTGGTCTGCTACTCTCTCCAGT.....AATGATCAACCCCTA	377
168	GluVal.....AsnIleProProGlyAlaLysValGlnPheGluLeuHI	182

378	CACGTGACAGATG.....GGCAACTGTGGAGAGAGGGTGAAGGATCC	421
182	styrGlnGlnMetLysTrpArgLysLeuGlySerTyrGluHisLysIleH	199
422	ACCTCACTCTGATTTCATTCAGGAAAAAGTTAGCTGAATATGGACCA	471
199	IsLeuGlnProGlyArgLeuAla.....LysHisLeuGluValAsnVal	213
472CAAGGTAGGGCATTTTCCATGATGGGCTCATCT	506
214	TrpIleValGluLeuGlnGlyMetArgPheLeuHis.....	225
507	ACGATGGGGAGTATTTCAGAGTACATAATATGATGAGAAATTCATTAT	556
226ValProAspThrPhe.....	230
557	CCAATGGAAGATACACAGCAGTAAGA...TGTTACGAGGATATTACTCGT	603
231	..GluGlyHisPheGlnGlyValProValIleSerLysGlyGlnLysLys	246
604	ACAAATGTAGTAAAGAAG.....TGTCAGGGAGGACGCTGTTACAC	644
247	SerHisValSerPheLysProThrValAlaGlnGlnArgLysCys.....	261
645	CAAAAGATGCACATTCATAAAGTAACAGGACTCTATGAAAAAGGATGTG	694
262	ProAsnCysThrTyrThrAlaValAspGly.....	271
685	AGTTTGTCTCCAATCCCGCCAGAGGAGGCTTCTATAATGTTTGA	744
271	271
745	CAACATGTTGATTCATATAGTTGAATTCGTACAGACAAAACCAACAA	794
272GluLeuValValMetTyrAspValAsnAr	281
795	AGAAGCTCCAAACAGCAAAATCAAAATCTCGAAGACACATGG	844
281	gGlu.....G	283
845	AAGTGATCCGTGATTCCTGAGGACTTTAAGAAACACCTCTATGACAACA	894
283	LuLysValGlyGluLeuGluValPheAsnGlyTyrPheValHisPhePhe	299
895	CAGCCACCAAAATCCACCTTCTCATTTGCTGCAGATTGGCAAAAGATGT	944
300	AlaProGluAsn.....LeuAspProIleProLysAsnII	311
945	GTGTTTAGTCTCTTGACAAATCTGGAAGCATGGCGACTGGTAACCGCTCA	994
311	eLeupPheValIleAspValSerGlySerMet...TrpGlyIleLysMetL	327
995	ATCGACTGAATCAACAGCCGACGCTTTTCCTG.....CTGCAGACAGTT	1038
327	ysGlnThrValGluAlaMetLysThrIleLeuAspAspLeuArgThrGlu	343
1039	GAGCTGGGCTCCTGGTGGGATGCTGACATTTGACAGTGTGCTGCCCATGT	1088
344	AspGlnPheSer.....ValValAspPheAsnHisAsnValArgTh	357
1089	A...CAAAAGTGAATCATACAGATAAAGTGGCAGTGCAGGGACACAC	1135
357	rTrpArgAsnAspLeuValSerAlaThrLysThrGlnIleThrAsp....	372
1136	TGCCCAAAAGATTA.....CCTGCAGCAGCTTCAGGAGGACGCTCCATC	1179
373	..AlaLysArgTyrIleGluLysIleGlnProSerGlyGlyThrAsnIle	388
1180	TGCAGCGGGCTTCGATCGCATTTTACTGTGATTAGGAAGAAA.....	1221
389	AsnGluAlaLeuArgAlaIlePheIleLeuAsnGluAlaSerAsnLe	405
1222TATCCAAGTGGATCTGAAATTTGCTGCTGACGGATG	1261

```
405 uGlyMetLeuAsnProAspSerValSerLeuIleValLeuValSerAspG 422
1262 GGAAGACAACTATAAGTGGGTCTTTAACGAGGTCTCAACAAAGTGGT 1311
422 ly...AspProThrVal.....GlyGluLeuLys..... 430
1312 GCCATATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAAGTAGA 1361
430 ..... 430
1362 GGAGCTGTCCAAATGACAGGAGTTTACAGACATATGCTTCAGATCAAG 1411
431 .....LeuSerLys.....IleGlnLysAsnValLysGlnAsnI 442
1412 TTCAGAACATGGCCCTCATTTGATGCTTTTGGG..... 1443
442 leGlnAspAsnIleSerLeuPheSerLeuGlyIleGlyPheAspValAsp 458
1444 .....GCCCTTTTCATCAGGAATGGAGCTCTCTCAGCG 1478
459 TyrAspPheLeuLysArgLeuSerAsnGluAsnArgGlyIleAlaGlnAr 475
1479 CTCATCCAGCTTGAGAGTAAGGATTAAACCTCCAGAACAGCCAGTGA 1528
475 gile.....TyrGlyAsnArgAspThrSerSerGlnLeuLysLysPheT 490
1529 TGAATGGCACAGTATCGTGGACAGCACCGTGGGAAAGGACACTTTCTT 1578
490 yrAsnGln.....ValSerThrProLeuLeuArgAsnValGlnPhe 503
1579 CTTATCACCTGGACAGCGACCTCCCAATCTCTCTGGGATCCCGAG 1628
504 .....AsnTyrProGlnAlaSerValThrAsp..... 512
1629 TGCAGACAGCAAGGTGGCTTTGTAGTGGACAAAAACACAAATGGCCT 1678
513 .....ValThrGlnAsnSerPheHisAsnT 521
1679 ACTCCAAATCCAGGATTTGCTTAAGTTGGCAGCTTGGAAATACAGTCTG 1728
521 yrPheGlyGlySerGluIleValValAlaGlyLysTyrAspProSerLys 537
1729 CAGCAGCTCACAACCTTGACCTGACCTGCTACGCTCCCGTGGCTCCAA 1778
538 LeuAlaGluValGlnSerIle...IleThrAlaThrSerThrAsnThrGI 553
1779 TGTACCTCTCCCTCCCAATTACA.....GTGACTT 1807
553 uLeuValLeuGluThrLeuSerGlnMetAspAspLeuGluAspPheLeuS 570
1808 CCAAAACGAACAGGACACAGCAATTTCCCGCCCTCTGGTAGTTAT 1857
570 erLysAspLysHisAlaAspProAsnPheThrLysLysLeuTrpAlaTyr 586
1858 GCAATATTCGCCAA.....GGAGCCTCCCAATTCTCAG 1892
587 LeuThrIleAsnGlnLeuLeuAlaGluArgSerLeuAlaPro..... 600
1893 GGCCAGTGTCCAGCCCTGATTGAATCAGTGAATGGAATAACAGTT.... 1938
601 .....ThrAlaAlaIleLysArgLysIleThrLysThrIleLeuG 614
1939 .....ACCTTGGAACTACTG 1953
614 lnMetSerLeuAspHisIleValThrProLeuThrAlaMetValIle 630
1954 GATAATGGACAGGTGCTGATGCTACTAAGGATGACGGTCTCTACTCAAG 2003
631 GluAsnGluAlaGlyAspGluArgMetLeuAlaAspSerPro..... 644
2004 GTATTTCAACAACCTATGACACAGGAATGGTAGATACAGTGTAAAAGTGGCGG 2053
```

```
644 ..... 644
2054 CTCTGGGAGAGTTAAAGCAGCAGCAGGAGAGTATACCCCGACAG... 2100
645 .....ProGlnAspHis 648
2101 .....AGTGAGCACTGTAC..... 2115
649 SerCysCysSerGlyAlaLeuTyrTyrGlyThrLysValAlaSerAlaSe 665
2116 ATACTGGCTGG..... 2127
665 rIleProSerTrpAlaSerProSerProThrProValMetAlaMetLeuA 682
2128 ..ATTGAGATGATGAATACAAATGAATCCACCAAGACCTGAAATTAAT 2175
682 laValGlyAlaAsnArgLeuGluSerThrProProHisValIleArg 698
2176 AAGGATGATGTTCAACAC.....AGCA 2198
699 ValGluAsnAspProHisPheIleIleTyrLeuProLysSerGlnLysAs 715
2199 AGTGTCTTTTCAGCAGAACATCTCGGAGGCTCATTTGTGGCTTCTCATG 2248
715 nIleCysPheAsnIleAspSerGlu..... 723
2249 TCCCAATGCTCCCATACCTGATCTCTTCCACCTGGCCAAATCCAGCAG 2298
724 .....ProGlyLysIleLeuSer 729
2299 CTGAAGCGGAAATTCACGGGGCAGCTCATTAATCTGACTTGGACAGC 2348
730 LeuValSerAspProGlnSerGlyIleLeuValAsn...GlyGlnLeuI 745
2349 TCCTGGGATGATTATGACCATGGAACAGCTCACAAAGTATATC..... 2391
745 eGlyAlaLysLysAlaGluAsnGlyLysLeuArgThrTyrPheGlyLysL 762
2392 .....ATTGCAATAGTACAAGT 2409
762 euGlyPheTyrPheGlnLysGluAspMetLysIleGluIleSerThrGlu 778
2410 ATTCTTGATCTCAGACACAGTTCATGATCTCTTCAAGTGAATACTAC 2459
779 AsnIleThrLeu.....IleAsnGlySer.....SerThrTh 789
2460 TGCTCTCTCCCAAGGAAGCC.....AACTCTGAGGAAGTCTTTT 2500
789 rSerLeuPheTrpSerAspThrAlaHisLeuGlyAsnGlnArgValLeuI 806
2501 TGTTT.....AAACCAGAAACATTACTTTTGAATAATGCCACAGATCTT 2544
806 leSerValLysLysGlyLysSerValThrLeuThrLeuAsnLysGluMet 822
2545 TTCATTGCTATTCCAGGCTGTTGATAGGTCGATCTGNAATCAGAAATATC 2594
823 PhePheSerVal...LeuLeuHisHisValTrpLysLysHisProVal... 837
2595 CAACATTGCACGAGTATCTTTGTTTATTCTCCACAGATCCGCCAGAGA 2644
838 AsnValAspPheLeuGlyIleTyrLeuProPro...ThrAsnLysPheS 853
2645 CACCTAGTCTGTATGAACAGCTCTGCTCCTTGT.....CCTAATATT 2685
853 erProSerAlaHisGlyLeuLeuGlyGlnPheMetAsnLysProAsnIle 869
2686 CATATCAACAGCACCATTCTCTGGC 2709
870 HisIlePheAsnGluArgProGly 877
seq_name: SwissProt_39:PTPD_HUMAN
seq_documentation_block:
```


1372 AAAATGACAGAGCTTTACAGACATATGCTTCAGATCAAGTTCAGAACAA 1421
898 ThrLeuThrValAsnAspAsnGlyLysAlaThrAspIleValGlnAsnSe 914
1422 TGGCCTC.....ATTGATGCTTTTGGG 1444
914 rGlyAlaAlaLeuGlnThrSerThrAlaAsnGlyIleGluIleSerGlyT 931
1445 CCCTTTTCATCAGGAATGAGCTGTCTCT.....CAGCGCTCCATC 1485
931 hrHisGlnTyrGlyThrPheSerIleSerGlyAsnLeuAlaThrAsnMet 947
1486 CAGCTTCAGAGTAGGAGTAACCTCCAGACAGCCAGCTGAGTAATGG 1535
948 LeuLeuGlnAsnGlyGly.....AsnLeuLeuValLeuAlaG 960
1536 CACAGTGATCGTGACAGCACCCTGGGAAGGACACATTTCTTTTATCA 1585
960 yThrGluAlaArgAspSerThrValGlyLys..... 970
1586 CTTGGCAACGCCCTCCCAATCTCTCTGGGATCCAGTGGACAG 1635
971GlyGly 972
1636 AAGCAAGTGGCTTTGAGTGGACAAAACACCAAAATGCCCTACCTCCA 1685
973 AlaMetGlnAsnLeuGlyGlnAspSerAlaThrLys.....Valas 986
1686 AATCCCGAGGATGCTAAGGTTGGCAGCTTGGAAATACAGTGCACAGCAA 1735
986 nSerGlyGlyGlnTyrThrLeuGlyArgSerLysAspGluPheGlnAlaL 1003
1736 GCTCACAAACCTTGACCTGACTGTC.....ACGTCCTGGCTGCC 1776
1003 euAlaArgAlaGluAspLeuGlnValAlaGlyGlyThrAlaIleValTyr 1019
1777 AATGCTACCTGCTCCCAATTACAGTACTTCCAAAACAGCAAGGACAC 1826
1020 AlaGlyThrLeuAlaAspAlaSerValSerGlyAlaThrGlySerLeuSe 1036
1827 CAGCAAAATCCCGACCTCTGCTAGTTTATGCAAAATTTCCGCAAGGAG 1876
1036 rLeuMetThrPro.....ArgAspAsnV 1044
1877 CTCCCAATCTCAGCGCCAGTGTACAGCCCTGATTGAATCAGTGAAT 1926
1044 alThrProValLysLeuGluGlyAlaValAlaGlyIleThrAspSerAla... 1059
1927 GGAATAACAGTTACCTTGGAACTACTGATAATGAGCAGGTGCTGATGC 1976
1060ThrLeuThrLeu.....GlyAsnGlyValAspTh 1069
1977 TACTAAGGATGAC.....GGTGTCTACT 1999
1069 rThrLeuAlaAspLeuThrAlaAlaSerArgGlySerValTrpLeuAsnS 1086
2000 CAAGGTATTTCAACATTATCAGCAAGTGTGATACATACAGTGTAAAGTC 2049
1086 erAsnAsnSerCysAlaGlyThrSerAsnCysGluTyrArgValAsnSer 1102
2050 CGGGCTCTG.....GGAGGAGTTAACGCCAGCAGGAGAGTGATACC 2093
1103 LeuLeuLeuAsnAspGlyAspValTyrLeuSerAlaGlnThrAlaAlaPr 1119
2094 CCAGCAGAGTGGACACTGTACATACCTGGCTGGATTGAGAATGAGAAA 2143
1119 oAlaThrThrAsnGlyIleTyr.....AsnThrLeuThrThrAsnGluL 1134
2144 TACAATGGAATCCACCAAGACCTGAATTAATAGGATGATGTTCAACAC 2193
1134 euSerGly.....SerGlyAsnPheTyrLeuHis 1143
2194 AAGCAAGTGTGTTTACAGACA.....ACATCCTC 2222

1144 ThrAsnValAlaGlySerArgGlyAspGlnLeuValValAsnAsnAl 1160
2223 GGGAGCTCAATT.....GTGGCTTCTGTATGCCCAATGCTCCCATAC 2266
1160 aThrGlyAsnPheLysIlePheValGlnAspThrGlyValSerProGlnS 1177
2267 CTGATCTCTCCACCTGGCCAATCACCCACCTGAGGCGGAATTCAC 2316
1177 erAsp.....AspAlaMetThrLeuValLysThr..... 1186
2317 GGGGCGAGTCTCATTAATCTGACTTGGACAGCTCCTGGGATGATTATGA 2366
1187 GlyGlyAspAlaSerPheThrLeuGlyAsnThrGlyGlyPheValas 1203
2367 CCATGGAACAGCTCACAAATATATCATTCGAATAAGTACAGTATCTTGT 2416
1203 pLeuGlyThr...TyrGluTyrValLeuLys.....SerA 1214
2417 ATCTCAGACAGCAAGTCAATGAATCTCTTCAAGTGAATACTACTGCTCTC 2466
1214 spGlyAsnSerAsnTrpAsnLeuThrAsnAspVal..... 1225
2467 ATCCAAAGGAAGCAACTCTGAGGAAGTCTTTTGTTTAAACACAGAAA 2516
1226LysProAsnProAspProIleProAsnProLysProAsp.. 1238
2517 CATTACTTTTGAATGGCACAGATCTTTTTCATTGCTATTTCAGGCTGTTG 2566
1239P 1239
2567 ATAAGTGCATCTGAATCAGAATATCCAACTGTCACAGATATCTTTG 2616
1239 roLysProAspProLys..... 1244
2617 TTTATTCTCCACAGACTCGCGCAGACACCTAGTCTGTGATGAAAGCTC 2666
1245ProAspProAsnProLysProAspProThrProAspProThrPr 1259
2667 TGCTCTCTGTCTAATATTCATATCAACACACCATTCCTCGGCATT 2712
1259 oThrProValProGluLysArgIleThrProSerThrAlaAlaVal 1274
seq_name: SwissProt_39:EAE_ECO11

seq_documentation_block:

ID EAE_ECO11 STANDARD; PRT; 935 AA.
AC O31000;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE INTIMIN (ATTACHING AND EFFACING PROTEIN) (EAE PROTEIN).
GN EAE OR EAEA.
OS Escherichia coli O111:H-.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=168927;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98187918; PubMed=9529069;
RA Voss E., Paton A.W., Manning P.A., Paton J.C.;
RT Molecular analysis of Shiga toxinigenic Escherichia coli O111:H-
RT proteins which react with sera from patients with hemolytic-uremic
RT syndrome.;
RL Infect. Immun. 66:1467-1472(1998).
CC - FUNCTION: NECESSARY FOR THE PRODUCTION OF ATTACHING AND EFFACING
CC - LESIONS ON TISSUE CULTURE CELLS. BELIEVED TO MEDIATE ADHERENCE.
CC - SUBCELLULAR LOCATION: OUTER SURFACE.
CC - SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.
CC - SIMILARITY: CONTAINS 1 LYSM REPEAT.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

```

DR ENBL; AF025311; AAC69247.1; -.
DR InterPro; IPR003344; Big_1.
DR InterPro; IPR003343; Big_2.
DR InterPro; IPR003535; Intimin.
DR InterPro; IPR002482; LysM.
DR Pfam; PF02369; Big_1; 2.
DR Pfam; PF02368; Big_2; 1.
DR Pfam; PF01476; LysM; 1.
DR PRINTS; PR01369; INTIMIN.
DR SMART; SM00257; LysM; 1.
DR Outer membrane; Virulence.
KW REPEAT 65 113
FT REPEAT 65 113
SQ SEQUENCE 935 AA; 101570 MW; 406E79CD07DEB11 CRC64;
LYSM.

```

```

alignment_scores:
  Quality: 133.00
  Ratio: 0.465
  Percent Similarity: 50.000
  Length: 572
  Gaps: 30
  Percent Identity: 20.979

alignment_block:
  US-09-049-696-18 x EAE_ECOL11 ..
  Align. seq 1/1 to: EAE_ECOL11 from: 1 to: 935

```

1096 GAACATCATACAGATAAACAGTGGC.....AGTGACAGGGA 1130
:::||||| ||||::: ::::: :::::
436 AspValValGlnArgAsnAsnIleLeuGluTyrlsLysGlnAs 452
:::||||| ::::: ||||
1131 CACACTCGCCAAAGAATTACCTGCAGCAGCTTCAGGAGGAGCTCCATCT 1180
| ||||::: ::::: ||||
452 pIleLeuSerLeuAsnIleProHisaspIleasnGlyThrGluHisSerT 469
1181 GCAGCGGGCTTCGATCGGCGATTACTGTGATTAGGAAGAAATATCCAAC 1230
::: ::::: ::::: |||||
469 hrGlnLysIleGln.....LeuIleValIlysSerLysrGlyLeu 482
1231 GATGATCTGAAATTTGTGCTGCTGCGGATGGGAAGACAACACTTAAG 1280
||| ::::: |||||
483 Asp.....ArgIleVal.....TrpAspSpSerAlaLeuar 493
1281 TGGGTGCTTTACGAGGTCACCAAGTGGTGCCTATCCACACAGTCG 1330
:::||||| ||| |||
493 g.....SerGlnGlyGlnIleGlnHisSer.... 502
1331 CTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGGAGCTG.....TCCAAA 1374
||| ::::: |||||::: ::::: |||
503GlySerGlnSerAlaGlnAspTryGlnAlaIleLeuProAlaTy 517
1375 ATGACAGCAGAGGTTTACAGACATATGCTTCAGATCAAGTTTCAGAACATGG 1424
::: ||||| ::::: |||
518 ValGlnGlyGlySerAsnIleTyrs..... 536
1425 CCTCATTTGATGCTTTTGGGCCCTTTTCATCAGGAATGGAGCTGTCTCTC 1474
::: ||| ::::: |||||
527ValThrAlaArgAlaTyraSpargAsnGlyAsn.....SerS 539
1475 AGCGCTCCATCCAGCTTGAGATTAAGGGATTAAACCTCCAGAACACCCAG 1524
::: ||||| ::::: |||||
539 erAsnAsnValGlnLeuThrIle.....ThrValLeuSerAsnGlyGln 553
1525 TGGATGAATGGCACAGTGTATCTGTGGACACGCGTGGGAAG..... 1566
::: ::::: ||||| ::::: |||
554 ValValASPGLnValcylValThrAspPheThrAlaaspLysThrSerAl 570
1567GACACTTTGTTCTTATCACCTGGGACACGCGCTCCCCAAA 1609
||| ::::: |||||::: |||||

570 aLysalaAspGlyThrGluAlaIleThrTyrThrAla..... 582

1610 TCCTTCTCTGGATCCAGTCAGACAGAAGGTCGCTTGTTGATGGAC 1659

583ThrValIysAsnGly.....ValThrGln 591

1660 AAAAACACCAAAATGGCTACCTCCAATCCCGACCATTCGTAAGTTGG 1709

592 AlaAsnValProValSerPheAsnIleValSerGlyThrAlaThrLeuGI 608

1710 CACTTGGAAATACACTCTGCAAGCAAGCTCACAACCTTCACCCCTGACTG 1759

608 yAlaAsnSerAlaThrAspAlaAsnGlyLysAla.....ThrV 622

1760 TCACGTCCCGTGGTCCAATGCTACCTGCTCCCAATTACAGTCAGCTTC 1809

622 alThrLeuIysSerSer.....ThrProGlyGlnValValSerAla 636

1810 AAAACACAAGGACACACCAAAATTCCTCCCGACCTCTGGTAGTTATGC 1859

637 LysThrAlaGluMetThrSerAlaLeuAsnAlaSerAlaValIlePheVa 653

1860 AAATATTCCCAAGGACCTCCCAATCTCAGGGCAGTGTCCACAGCCC 1909

653 lGluGlnThrLysAlaSerIleThrGluIleLysAlaAspLysThrThra 670

1910 TCATTCAATCAGTGAAT...GGA AAAACAGTTACTCTGGAACTACTCGAT 1956

670 lavalAlaAsnGlyAsnAspAlaValThrThrValIysValMetLys 686

1957 AATGGACAGGTGTGATGCTACTAAGGATGACGGTGTCTACTCAAGGTA 2006

687 GluGlyGlnProValHisGly.....HisSerValAl 697

2007 TTTCACACTTAT.....GACACGAATGGTAGATACAGTCTAAAAGTCC 2050

697 aPheThrThrAsnPheGlyMetPheAsnGlyLysSerGlnThrGlnAsnA 714

2051 GGCTCTCGGAGGAGTAAACCCAGCAGGAGAGTGATACCCACAGCAG 2100

714 laThrThrGlySerAspGlyArgAlaThrIleThrLeuThrSerSerSer 730

2101 AGTGGACAGCTGTACATACCTGGCTGGATGAGAAATGATGAATCAATG 2150

731 AlaGlyLysAlaThrValSerAlaThrValSerGlyGlyAsn..... 744

2151 GAATCCACCAAGCTGAAAT.....A 2173

745 AspValLysAlaProGluValThrPheAspGlyLeuLysIleAspA 761

2174 ATAAGGATGATGTTCAACACAAGCAAGCTGT..... 2205

761 snLysValAspIleLeuGlyLysAsnValThrGlyAspLeuProAsnIle 777

2206TTCAGAGAATCTCTCGGAGGCTCATTTGT 2237

778 TrpLeuGlnTyrGlyGlnPheLysLeuLysValSerGlyGlyAsnGlyTh 794

2238 GCCTTCT...GATGTCCCAATGTCCCATACCTGATCTCTTCCCACTG 2284

794 rTyrSerTrpHisSerGluAsnThrAsnIleAlaThrValAspGluSerG 811

2285 GCCAATCACCAGCTGAAGCGGAAATTCACGGGGCAGCTCTCAATTA 2334

811 lLysValThr...LeuLysGlyLys.....GlyThrAlaValIleAsn 824

2335 CTGACTTGGACGCTCTGGGAGTGATTATGACCATGGA.....AC 2375

825 ValThr.....SerGlyAspLysGlnThrValSerTyrThrIleY 838

2375 AGCTCACAAGTATATCAVTCGAATAAGTACAAGTATTCATTGATCTCAG 2425

838 salaProAsnTyrMetIleArgValGly..... 847


```
1812 .....HisAsnGlyLysLeuProThrArgAlaMetAspPheAsnGlyIle 1826
1866 GGTGCTGATGCTACT .....AAGGTGACGGTGTCTAC ..... 1998
1827 LeuThrIleArgAsnValGlnLeuSerAspAlaGlyThrTyrValCysTh 1843
1999 .....TCAAGGTATTTCACAACTTATGACACGAATGTGTAGATACAGTGTAA 2044
1843 rGlySerAsnMetPheAlaMet .....AspGlnGlyThrAlaThrLeuH 1858
2045 AACTGGCGGCTCTGGGAGGAGTTAACCCAGCCAGACGGAGAGTGATACCC 2094
1858 isValGlnAla ..... 1861
2095 CAGCAGAGTGGAGCACTGACTACATACCTGGCTGGTGTGAGAAATGATGAAAT 2144
1862 .....SerGlyThrLeuSerAlaPro .....ValVa 1870
2145 ACAATGGAATCCACCAAGACCTGAAATTAATAGGATGATGTTCAACACA 2194
1870 lserIleHisProGlnLeu .....ThrValGlnProG 1882
2195 AGCAAGTGTGTTTCACGAGAACATCTCGGAGGCTCATTTGTGGCTTCT 2244
1882 lyGlnLeuAlaGluPheArgCysSerAlaThrGlySer ..... 1894
2245 GATGTCCCAAAATGCTCCCATACCTGATCTC .....TTCCACCC 2282
1895 .....ProThrProThrLeuGluTrpThrGlyGlyProG 1906
2283 TGCCCAATCACCAGCTGAAGCGGAAATTCACGGGGGAGTCTCATTA 2332
1906 yGlyGlnLeu...ProAlaLysAlaGlnIleHisGlyGly...IleLeuA 1921
2333 ATCTGACTTGGACAGCTCCGCGGATCATTTATGAC ..... 2367
1921 rGleuProAlaValGluProThrAspGlnAlaGlnTyrLeuCysArgAla 1937
2368 CATGGAACAGCTCACAGTATATCATTCGAAATAGTACAGTATCTTGA 2417
1938 HisSerSerAlaGlyGlnGlnValAlaArg .....AlaValLeuH 1951
2418 TCTCAGACAGCAAGTTCNAATGAATCTCTCAAGTGAAT ..... 2454
1951 sValHisGlyGlyGlyProArgValGlnValSerProGluArgThrG 1968
2455 .....ACTACTGCTCTC ..... 2466
1968 lnValHisAlaGlyArgThrValArgLeuTyrCysArgAlaAlaGlyVal 1984
2467 .....ATCCCAAGGA 2477
1985 ProSerAlaThrIleThrTrpArgLysGluGlyGlySerLeuProProG 2001
2478 AGCCAACTCTGAGAA .....GTCCTTTTGTAAACACAGAAACA 2518
2001 nAlaArgSerGluArgThrAspIleAlaThrLeuLeuProAlaIle 2018
2519 TTACTTTTGAATGGCAGAGATCTTTTCATTCTATTCAGGCTGTGTAT 2568
2018 hrThrAlaAspAlaGlyPheTyrLeuCysValAlaThrSerProAlaGly 2034
2569 AAGGTGATCTGAAATCAGAAATATCCAACTTGCAGGAGTATCTTTGTT 2618
2035 ThrAlaGlnAlaArgIleGlnValValLeuSerAlaSerAspAlaSe 2051
2619 TATTCCTCCACAGACTCCGCCAGACACCTAGTCTCTGTATGAACGTCTG 2668
2051 rGlnProPro...ValLysIleGluSerSerProSerValThrGluG 2067
2669 CTCCTTGTCTTAATATATATATCAACAGACGACCATTCCTGCATTCACATT 2718
```

```
2067 ly .....GlnThrLeuAspLeuAsnCysValValAlaGlySerAlaHis 2081
2719 TTAATAATATTATGTGG 2733
2082 AlaGlnValThrTrp 2086

seq_name: SwissProt_39:PMPE_CHLMU
seq_documentation_block:
ID PMPE_CHLMU STANDARD: PRT: 976 AA.
AC Q9PL47:
DC 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE OUTER MEMBRANE PROTEIN PMPE PRECURSOR (POLYMORPHIC MEMBRANE
DE PROTEIN E).
GN PMPE OR TC0261.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT Pneumoniae AR39*.
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -|- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC -|- (POTENTIAL).
CC -|- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE002293; AAF39130.1; -
DR TIGR; TC0261; -
DR InterPro; IPR003357; OMP.
DR Pfam; PF02385; OMP; 1.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL : 1 18 POTENTIAL..
FT CHAIN : 19 976 PROBABLE OUTER MEMBRANE PROTEIN PMPE.
SQ SEQUENCE 976 AA; 105964 MW; 0BEA56158E3CEE9 CRC64;

alignment_scores:
Quality: 131.50 Length: 861
Ratio: 0.353 Gaps: 39
Percent Similarity: 43.206 Percent Identity: 18.118

alignment_block:
US-09-049-696-18 x PMPE_CHLMU ..
Align seg 1/1 to: PMPE_CHLMU from: 1 to: 976

238 TTTTATTTCAAAAATGTTGCCATTTTTCATTCCTGAAACATGGAAGACAAA 287
239 ||| ||||| ||| ||||| ||| ||||| |||
240 PhcCysPheLysAsn .....LeuThrProGluSerGlyGlyValI 108
241
285 GGCTGACTATGTGAGA .....CCAAACTTGACCTACAAAATGCTG 331
109 eGlySerProThrGlnAsnThrProThrIleLysIleHisAsnThrIleG 125
332 ATGTTCTGTTGCTGACTCT .....ACT 354
```

```

333 snHisThrGlyClyAlaIleTyrIleThrGlyThrGluThrSer..... 347
1207 GTGATTAGGAAGAATAATCCAACTGATGGATCGTCAAATTTGTGCTGTCAC 1256
      ||| ||||| ||||| ||||| ||||| |||||
348 .....LysIleSerAlaAspHisAlaIleIlePhe..... 358
1257 GCATGGGAAACACACACTATAAGCTGGTGCTTTAAACGAGGTCAACAACAA 1306
      ||||| ||||| ||||| ||||| ||||| |||||
359 .....AspAsnAsnIleSerAla.....AsnAlaThrAsnAlaA 370
1307 GTGGTGCCATCATCCACACA..... 1326
      ::::: |||||
370 spClySerSerSerAsnThrAsnProProHisArgAsnAlaIleThrMet 386
1327 .....GTCCCTTTGGGCCCTCTCGACGCTCAAGNACT 1358
      :: ||||| ||||| ||||| ||||| |||||
387 AspAsnSerAlaGlyGlyLeuGlyAlaGlyLysSerGln..... 401
1359 AGAGCAGCTGTCCAATAATGACAGGAGTTTTACACACATATGCTTCAGATC 1408
      :||| ||||| ||||| ||||| |||||
402 .....AsnLeuIlePheTyrAspProIleG 410
1409 AAGTTTCAGAACAAATGGCCTCATTTGATGCTTTGGGCGCCTTTTCATCAGGA 1458
      ||||| ||||| ||||| ||||| ||||| |||||
410 InValThrAsnAlaGlyValThrValAspPheAsnLysAspAlaSerGln 426
1459 AATGAGCTGTCTCTCACGCGCTCCATCCAGCTTGAGAGTAAGGATTAAC 1508
      :||| ||||| ||||| ||||| |||||
427 ThrGlyCysValVal..... 431
1509 CTTCCAGAACACCAGTAGTGGTGAATGGCAGACGATGATCGTGGACAGCACCG 1558
      :||| ||||| ||||| ||||| ||||| |||||
432 .....PheSerGlyAlaThrValLeuSerAlaAspI 442
1559 TGGGAAGACACTTTGTCTTATCACTCGGACAAAGCGCAGCCTCCCCAA 1608
      :||||| :||| |||||
442 IleSerGlnAlaAsnLeu.....GlnThrLysThrProAlaThr 454
1609 ATCTCTCTCTGGATCCCAGTGGACAGCAAGCTGGCTTTGTAGTGGGA 1658
      :||| ||||| ||||| ||||| ||||| |||||
455 LeuThrLeu.....SerHisGlyLeuLeuCysIleGl 465
1659 CAATAACACCAAAATGGGC.....TACTCCAAATCCAGGCATTTG 1699
      :||| ||||| ||||| ||||| ||||| |||||
465 uAspArgAlaGlnLeuThrValAsnAsnPheThrGlnThrGlyGlyIleV 492
1700 CTAAAGTTGGC.....ACTTGGAAATACAGTCTCCAA 1731
      :||| ||||| ||||| ||||| |||||
482 AlaLeuGlyAsnGlyAlaValLeuSerSerTyrGlnHisSerThrThr 498
1732 GCAAGTCAAACTTGACCTGACTCTACGTCAGTCCCGTGGTCC..... 1776
      :||| ||||| ||||| ||||| ||||| ||||| |||||
499 AspAlaThrGlnThrProProThrThrThrThrThrThrThrThrThrThr 515
1777 .....AATGCTACCTCGCTCCCAATTAAGTACGTACTTCCAAAA 1813
      :||| ||||| ||||| ||||| ||||| |||||
515 rLeuAsnHisIleGlyLeuAsnLeuProSerIleLeu..... 527
1814 CGAACAAGACACCAGCAAAATCCCCAGCCCTCTGGTAGTTTATGCAAAAT 1863
      ||||| ||||| ||||| ||||| ||||| |||||
528 ....LysAspGlyAlaGluMetProLeuLeuTrpValGluProIleSer 542
1864 AITGCCAAGGAGCCTCCCCAATTCTCAGGCCAGTGTCCACAGCCCTGAT 1913
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
543 ThrThrGlnGly...AsnThrThrThrThrThrThrThrThrThrThrThr 558
1914 TGAATCAGTGAATGGAAAACAGTTTACCTTGGCACTACTGGTGAATATGAG 1963
      :||| ||||| ||||| ||||| ||||| ||||| |||||
558 rPheSerLeuAsnGly.....AlaThrLeuSerLeuIleAsp..... 570
1964 CAGGTGCTGATGCTACTAAGNATGACGGTGTCTACTCAAGGTATTTTACA 2013
      :||| ||||| ||||| ||||| ||||| ||||| |||||
571 .....GluAspGly...AsnSerProTyrGluAsn 579

```


190 YAlaThrThrAsnAsnIleLysLeuThrGlyIleThrLeuSerThrG 207
1598 AGCTCCCAATCCCTCTCTGGATCCCGATGACAGCAAGCAAGTGGC 1647
207 luAlaAlaThr.....AspThrGlyThrAsnProAlaSer 219
1648 TTT.....CTAGTGGCAAAACACCAAAATGCCTACCTCCAAATCCC 1691
220 IleuLysValThrThrAspAsnGlyAsnAspTyrTyrAlaLysIleTh 236
1692 AGCATGCTAAGTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCAC 1741
235 rGlyGlyAspAsnAspGlyLysTyrTyrAlaValThrValAlaAsnAspG 253
1742 AAACCTTGACCTGACTGCTACGCTCCGCTGCTCCCAATGCTACCTGCGCT 1791
253 lyThrValThrMet....AlaThrGlyAlaThrAlaAsnAlaThrValThr 268
1792 CCAATTACAGTGACTTCCAAACGAACAGCACACCAAGCAAAATTCCTCCAG 1841
269 AspAlaAsnThrThrLysAlaThr..... 276
1842 CCCTCTGTAGTTTATGCAATATTCGCAAGAGCCTCCCAATTCCTCA 1891
277ThrIleThrSerGlyThrProVal...G 286
1893 GGGCCAGTGTACAGCCCTGATTGAATCAGTGAATGGAACAGTGTACC 1941
286 lIleAspAsnThrAlaGlySerAlaThrAlaAsnLeuGlyAlaValSer 302
1942 TTGGAACCTACTGGATAATGGAGCAGCTGCTGATCTACTAAGGATGACGG 1991
303 LeuValLysLeuGlnAspSerLysGlyAsnAspThrAspThr..... 316
1992 TGCTACTCAAGGTATTTCACAACTTATGACACGAATGGTAGA...TACA 2038
317TyrAlaLeuLysAspThrAsnGlyAsnLeuTyrA 328
2039 GTGTAAGAGTGGGCTCTGGAGGAGTTACGGCAGCAGCAGGAGGTG 2088
328 laAlaAspValAsnGluThrThrGlyAlaValSerValLysThrIleThr 344
2089 ATACCCACGACAGTGGAGCAGCTGTACATACCTGCTGCTGATGAGAATGA 2138
345 TyrThrAspSerSerGlyAlaAlaSer..... 353
2139 TGAATACATGGAATCCACCGACCTGAAATTAATAGGATGATGTC 2188
354SerProThrAlaValLysLeuGlyGlyAspAsp..... 364
2189 AACACAAGCAAGTGTGTTTCAGCAGAATCTCCTGGGAGGCTCATTTGTG 2238
365GlyLysThrGlu 368
2239 GGTCTGTGATCCCAATGCTCCCATACCTGATCTCTCATTAAT... 2334
375 sThrTyrAsp....SerAlaAspLeuAsnGlyGlyAsnLeuGlnThrGlyL 391
2335CTGACTTGGACAGCTCCTGGGATGATTAT 2364
391 euThrAlaGlyGlyGluAlaLeuThrAlaValAlaAsnGlyLysThrThr 407
2365 GACCATGGAAACACCTCACAAGTATATCATTCGAATAGTACAGTATCT 2414
408 AspProLeuLysAla.....LeuAspAlaIleAl 418
2415 TGATCTCAGACAGCAAGTTCAAATGAATCTCTCAAGTGAATACTACTGCTC 2464
418 aSerVal....AspLysPheArgSerLeuGlyAlaValGlnAsnArgL 434

DR Pfam: PF00669; Flagellin_N; 1.
DR PRINTS: PR00207; FLAGELLIN.
DR ProDom: PD000316; Flagellin_C; 1.
KW Flagella; Complete proteome.
FT INT_MET 0
FT CONFLICT 284 284 P -> L (IN REF. 2).
SQ SEQUENCE 497 AA; 51163 MW; 24B5419C21C7B4E8 CRC64;

alignment_scores:
Quality: 131.00 Length: 584
Ratio: 0.468 Gaps: 26
Percent Similarity: 47.945 Percent Identity: 20.377

alignment_block:
US-09-049-696-18 x FLIC_ECOLI ..
Align seg 1/1 to: FLIC_ECOLI from: 1 to: 497

970 AGCATGGCAGCTGGTAACGCGCTCAATCGACTGAATCAAGCAGCGCAGCT 1019
10 SerLeuIleThrGlnAsnAsnIleAsnLys...AsnGlnSerAlaLeuSe 25
1020 TTTCTCTGCTGCAGACAGTTCAGCTGGGCTCCTGGTGGGTGGTGCAT 1069
25 rSerSerIleGluArgLeuSerSerGly..... 34
1070 TTGACAGTGTGCCCATGTATCAAGTGAATCATACAGATAAAGTGGC 1119
35LeuArgIleAsnSerAla 40
1120 AGTGACAGGACACATCGCGCAAAAGATTACCTGCAGCAGCTTCAGGAGG 1169
41 LysAsp...AspAlaAlaGlyGln.....AlaIleAlaAsnArgPh 53
1170 GAGCTCCATCTGACGGGGCTTCGATCGCGCATTTACTGTCAATTAGGA 1219
53 eThrSerAsnIleLysGlyLeuThrGlnAla.....AlaA 65
1220 AATATCAACTGATGATCGAATGTGCTGCTGACGAGTGGGAGGAC 1269
65 rGlnAlaAsnAspGlyLysSerValAlaGlnThrThrGluGlyAlaLeu 81
1270 AACACTATAAGTGGTCTTTACAGAGGTCAACAA..... 1305
82 SerGluIleAsnAsnAsnLeuGlnArgValArgLeuThrValGlnAl 98
1306AGTGTGCCATCATCCACAGCTCGCTTTGGGCGCTCTG 1345
98 aThrThrGlyThrAsnSerGluSerAspLeuSerSerIleGlnAspGluI 115
1346 CAGCTCAAGAACTAGAGAGCTGTCCAAATGACAGGAGGTTTACAGACA 1395
115 leLysSerArgLeuAspLysIleAspArgValSerGly..... 127
1396 TATGCTTCAGATCACTTCAGCAACATGGC.....CTCATGATGCTTT 1439
128GlnThrGlnPheAsnGlyValAsnValLeuAlaLysAs 140
1440 TGGGGCCCTTTTCAGGAAATGAGCTGTCTCTCAGCGCTCCATC.... 1485
140 nGlySerMetLysIleGlnValGlyAlaAsnAspAsnGlnThrIleThrI 157
1486CAGCTTGAAGTGAAGGATTAAACCTCCAGACAGCCAG 1524
157 leAspLeuLysGlnIleAspAlaLysThrLeuGlyLeuAspGlyPheSer 173
1525 TGGATGAATGACAGTGCCTGGACAC..... 1554
174 ValLysAsnAsnAspThrValThrThrSerAlaProValThrAlaPheG 190
1555ACCGTGGAAAGGACACTTTCTTTCATCTACCTGGACAGCG 1597

```
2465 TCATCCCAAGGAGCAACTCTGAGGAGTCTTTTGGTTTAAACACAGAA 2514
|||
434 euAspSerAlaValThrAsnLeu.....Asn 442
2515 AACATTACTTTGAAATGCGACAGATCTTTTCATTCTATTGAGGCTGT 2564
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
443 AsnThrThr.....ThrAsnLeuSerGluAlaGlnSerArgIi 455
2565 TGATAAGTCTGCTGAAATCAAGAAATATCAACATTCACGAGTATCTT 2614
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
455 eGlnAspAlaAspTyrAlaThrGluValSerAsnMetSerLysAlaGlnI 472
2615 TG 2616
472 le 472
```

seq_name: SwissProt_39:FLIC_SALMU

seq_documentation_block:

```
AC P06177; STANDARD; PRT; 504 AA.
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FLAGELLIN (PHASE-1-D FLAGELLIN).
GN FLIC.
OS Salmonella muenchen.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI-TaxID=596;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 8388;
RX MEDLINE=86143833; PubMed=4093985;
RA Wei L.-N., Joys T.M.;
RT "Covalent structure of three phase-1 flagellar filament proteins of
RT Salmonella."; 186:791-803(1985).
RL J. Mol. Biol.
CC -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -!- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE
CC BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLA, TERMED
CC PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

alignment_scores:

```
Quality: 130.50 Length: 457
Ratio: 0.565 Gaps: 20
Percent Similarity: 50.547 Percent Identity: 21.882
```

alignment_block:

```
US-09-049-696-18'x FLIC_SALMU
Align seg 1/1 to: FLIC_SALMU from: 1 to: 504
970 AGCATGGCGAGCTGGTAACCGCTCAATCGACTGAATCAAGCAGGCCAGCT 1019
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10 SerLeuLeuThrGlnAsnAsnLeuAsnLysSerGlnSerAla..... 23
1020 TTTCTCTCTGCACACAGTTCAGCTGGGGTCTCTGGTTCGGTGGTGCAT 1069
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
24 .....LeuGly..... 25
1070 TTGACAGTGTGCTCCCTGTACAAAGTGAACATCATACAGATAAAGTGGC 1119
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
26 .....ThrAlaIleGluArgLeuSerSerGlyLeuArgIleAsnSerAla 40
1120 AGTGACAGGGACACATCGCGCAAAAGATTACCTGCAGCAGCTTCAGGAGG 1169
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
41 LysAsp...AspAlaAlaGlyGln.....AlaIleAlaAsnArgPh 53
1170 GACGTCCATCTGCAGCGGGTTCGATCGCATCTTACTGTGATTAGGAAGA 1219
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
53 eThrAlaAsnIleLysGlyLeuThrGlnAla.....SerA 65
1220 AATATCCAAGTGTGATCTGAATTTGCTGCTGCTGACGGATGGGAAGAC 1269
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 rgAsnAlaAsnAspGlyIleSerIleAlaGlnThrThrGluGlyAlaLeu 81
1270 AACACTATAAGTGGTCTTTTACGAGGTCAACAA..... 1305
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
82 AsnGluIleAsnAsnLeuGlnArgValArgGluLeuAlaValGlnSe 98
1306 .....AGTGGTCCCATCATCCACAGTCGCTTTGGGGCCCTCTG 1345
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98 rAlaAsnGlyThrAsnSerGlnSerAspLeuAspSerIleGlnAlaGluI 115
1346 CAGCTCAAGACTAGAGGAGCTGTCCAAATATGACAGCA..... 1383
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
115 leThrGlnArgLeuAsnGluIleAspArgValSerGlyGlnThrGlnPhe 131
1384 ...GGTTTACAGACATATGCTTCAGATCAAGTTCAGAACAAATGACAGCA 1430
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
132 AsnGlyValLysValLeuAla.....GlnAspAsnThrLeuTh 144
1431 TGATGCTTTTGGGGCCCTTTTCATCAGAAATGGAGCTGTCTCTCAGCGCT 1480
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
144 rIleGlnValGlyAlaAsnAspGlyGluThrIleAspIleAspLeuLysG 161
1481 CCATCCAGCTTGAGAGTAAGGGATTAACTCCACAGACAGCCAG..... 1524
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
161 lulleSerSerLysThrLeuGlyLeuAspLysLeuAsnValGlnAspAla 177
1525 .....TGGATGAATGCGCAGTGTGTCGACAGCAGCCG..... 1560
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
170 TyrThrProLysGluThrAlaValThrValAspLysThrThrTyrLysAs 194
1561 .GGAAGGACACTTTGTTTCTTATCATCTGGACACAGCCAGCTCCCAAAA 1609
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
194 nGlyThrAspThrIle.....ThrAlaGlnSerAsnThrA 206
1610 TCCTTCTCTGGATCCCGAGTGGACAGAGCAGGCTGCTTTGTAGTGGAC 1659
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
206 spIleGlnThrAlaIleGlyGlyGlyAlaThrGlyValThrGlyAlaAsp 222
1660 AAAAACACCAAAA.....ATGGGCTACCTCAAAATCCAGGCATTGCTAA 1703
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
223 IleLysPheLysAspGlyGlnTyrTyrLeuAspValLysGlyGlyAlaSe 239
1704 GGTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAACCTTGACCC 1753
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
239 rAlaGlyValTyrLysAlaThrTyrAspGluThrThrLysLysValAsnI 256
1754 TGACTGTACAGTCCCGTGGTCCCAATGCTACCTGCCT..... 1791
```

```

::      :::::|||||:      ::|||
256 leAspThrThrAspLysThrProLeuAlaThrAlaGluAlaThrAlaLe 272
1792 .....:.....:CCAATTACAGTGACTTCCAAAC 1814
273 ArgGlyThrAlaThrIleThrHisAsnGlnIleAlaGluValThrLysGI 289
1815 GAACAGGACACAGCAGCAAAATCCCCAGCCCTGTGTAGTTATGCAATA 1864
289 uGlyValAspThrThrValAlaAlaGlnLeuAlaA...AlaGlyV 305
1865 TTCGCCAGGACGCTCCCAATCTCAGGCCAGTGTCACAGCCCTGATT 1914
305 al...ThrGlyAlaAspLysAspAsnThrSerLeuValLysLeuSerPhe 320
1915 GAATCAGTGAATGGAAACACGTTTACCTTGGAACTACTGGATAATGGAGC 1964
321 GluAspLysAsnGlyLysValle.....AspGlyGI 331
1965 AGGTGCTGATGCTACTAAGATGACGGTGTCTACTCAAGGTATTTCAAA 2014
331 yTyraValLysMetGlyAsp.....PheTyraAlaAlaT 344
2015 CTTATGAC.....ACGAAT 2028
344 hrTyraAspGluLysGlnValGlnLeuLeuLeuAsnHisTyrThrAsp 360
2029 GGTAGATACAGTGTAAAGTCGGGGCTCTG.....GGAGGAGTTAAAGC 2072
361 GlyAlaGlyValLeuGlnThrGlyAlaValLysPheGlyGlyAlaAsn... 376
2073 AGCCAGCAGGAGTATACCCAGCAGAGTGAGCAGTGTACATACCTG 2122
377 .GlyLysSerGluValValThrAlaThrValGlyLysThrTyrLeuAla 393
2123 GCTGGATTGAGATGATGAATACATGAAATCCACCAAGACCTGAAATT 2172
393 erAspLeuAspLysHisAsnPheArgThrGlyGlyCluLeuLysGluVal 409
2173 AATAAGAGTATGTTCAACAC 2193
410 AsnThrAspLysThrGluAsn 416
seq_name: SwissProt_39:RPB1_DROME

seq_documentation_block:
ID RPB1_DROME STANDARD; PRT; 1887 AA.
AC P04052; Q9VYX6;
DT 01-NOV-1986 (Rel. 03, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6).
GN RPII215 OR CG1554.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89218930; PubMed=2496296;
RA Jost R.S., Weeks J.R., Zehring W.A., Greenleaf A.L.;
RT "Analysis of the gene encoding the largest subunit of RNA polymerase
RL II in Drosophila.";
RL Mol. Gen. Genet. 215:266-275(1989).
[2]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RA Adams W.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,

```

```

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Piankocch C., Batowin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Besley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Buftis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA deRablos B.; Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodak A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
RA Hostin D., Harvey D., Helman I.J., Hernandez J.R., Houck J.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[3]
RP SEQUENCE OF 1-27 FROM N.A.
RX MEDLINE=87089662; PubMed=3025586;
RA Searles L.L., Greenleaf A.L., Kemp W.E., Voelker R.A.;
RT "Sites of P element insertion and structures of P element deletions in
the 5' region of Drosophila melanogaster RpiI215.";
RL Mol. Cell. Biol. 6:3312-3319(1986).
[4]
RP SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=85282618; PubMed=2992806;
RA Biggs J., Searles L.L., Greenleaf A.L.;
RT "Structure of the eukaryotic transcription apparatus: features of the
gene for the largest subunit of Drosophila RNA polymerase II.";
RL Cell 42:611-621(1985).
[5]
RP SEQUENCE OF 1441-1887 FROM N.A.
RX MEDLINE=88094402; PubMed=3122024;
RA Allison L.A., Wong J.K.-C., Fitzpatrick V.D., Moyle M., Ingles C.J.;
RT "The C-terminal domain of the largest subunit of RNA polymerase II of
Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a
conserved structure with an essential function.";
RL Mol. Cell. Biol. 8:321-329(1988).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
CC RNA(N).
CC -1- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM TEN TO
CC TWELVE DIFFERENT POLYPEPTIDES. THE 215 KDA POLYPEPTIDE IS THE
CC LARGEST COMPONENT OF RNA POLYMERASE II.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC THE PHOSPHORYLATION ACTIVATES POL2.
CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND TRNA GENES.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.

```


1176 CATCTCAGCGGCTTCGATCGCATTTACTGTG.....A 1210
1177 |||||:|||||
971 uproCysAsnLeuGlnArgMetIleThrPasnValGlnLysIlePheHisI 988
1211 TTAGGAAGAAATATCCAACTGATGATCGAAAT..... 1245
1212 |||||:|||||
988 leAsnLysArgLeuProThrAspLeuSerProIleArgValIleLysGly 1004
1246GTGCTGCTGACGGATGGGGAAGACAA 1271
1005 valLysThrLeuLeuGluArgCysValIleValThrGlyAsnAspArgII 1021
1272 CACTATAAGTGGGTGCTTTAAAGAGGTAAACAAAGTGGTGCATCATCC 1321
1021 eSerLysGlnAlaAsnGluAsnAlaThrLeuLeuPheGlnCysLeuIleA 1038
1322 ACACATCGCTTTGGGGCCCTCGCAGCTCAGAACTA..... 1359
1038 r9SerThrLeuCysThrLysTyrValSerGluGluPheArgLeuSerThr 1054
1360 GAGGAGCTGTCCAAATGACAGAGGTTCACAGACATATCCTTCAGATCA 1409
1055 GluAlaPheGluThrLeuValGluIleGluThrArgPheGlnGlnAl 1071
1410 AGTTCAGAACAAATGGC...CTCATTTGATGCTTTTGGGGCCCTTTCATCAG 1456
1071 aGlnAlaAsnProGlyGluMetValGlyAlaLeuAlaAlaGlnSerLeuG 1088
1457 GAAATGGAGCT.....GTC 1470
1088 lyGluProAlaThrGlnMetThrLeuAsnThrPheHisPheAlaGlyVal 1104
1471 TCTCAGCGCTCCATCCAGCTT..... 1491
1105 SerSerLysAsnValThrLeuGlyValProArgLeuLysGluIleLeAs 1121
1492GAGAGTAAGGATTAAACCTCCAGACAGCCAGT 1525
1121 nIleSerLysLysProLysAlaProSerLeuThrVal.....P 1134
1526 GGATGAATGACACAGTCTGTCGACAGCACCGTGGGAAAGGACACTTTG 1575
1134 heLeuThrGlyGlyAlaAlaArgAspAlaGluLysAlaLysAsnValLeu 1150
1576 TTCTTTATCACCTGGCAACG.....CAGCCTCCCAAAATCCT 1613
1151 CysArgLeuGluHisThrThrLeuArgLysValThrAlaAsnThrAlaI 1167
1614 TCTCTGGATCCCACTGGACAGAGCAAGTGGCTTTGTAGTGGACAAAA 1663
1167 eTyrTyrAspProAspProGlnArgThr.....ValIleSerGluA 1181
1664 ACACCAAAATG.....GCCTACCTCCAAATCCCGAGCAT.....GCT 1701
1181 spGlnGluPheValAsnValTyrTyrGluMetProAspPheAspProThr 1197
1702 AAGTTGGCACTTGGAAATACAGTCTGCAACGAGCTCACAACCTTCAC 1751
1198 ArgIleSerProThrLeuLeuArgIleGluLeuAspArgLysArgMet.. 1213
1752 CTGACTGTACGTCCTCGGTGGTCCCAATGTACCTGCCCTCCCAATTACAG 1801
1214ThrAspLysLysLeuThrM 1220
1802 TGACTTCCAAACGACAGACACAGCAATTCCTCCAGCCCTCTCGTA 1851
1220 etGluGlnIleAlaGluLysIleAsnValGlyPheGlyGluAspLeuAsn 1236
1852 GTTTATGCAATATTCGCCAAGGAGCTCCCAATTCCTCAGGGCCAGTGT 1901
1237 CysIlePheAsnAspAspAsnAlaAspLysLeuValLeuArgIleArgII 1253

1902 CACAGCCCTGATTGAATCAGTGAATGAAAAACAGTTTACCTTGGAACTAC 1951
1253 eMetAsnAsnGluGluAsn.....LysPheG 1262
1952 TGGATAATGAGCAGAGGTGCTGATCTACTAAGGATGACGGTCTACTCA 2001
1262 |||||:|||||
1262 lNaspGluAspGluAlaValAspLysMetGluAspAsp...MetPheLeu 1277
2002 AGGTATTTCACAACTTATGACACGAATGGTAGATACAGTGTAAAGTGCG 2051
1278 ArgCysIleGluAlaAsnMetLeuSerAspMetThrLeuGlnGlyIleGI 1294
2052 GCCTCTGGAGGAGCTT.....AACGCAGCCAGACGGA 2083
1294 uAlaIleGlyLysValTyrMetHisLeuProGlnThrAspSerLysLysA 1311
2084 CAGTGATACCCAGCAGAGTGGAGCAGCTGTACATACCTGGC...TGG... 2127
1311 rGileValIleThrGluThrGlyGluPheLysAlaIleGlyGluThrLeu 1327
2128 ATTGAATGATGAATACAAATGAATGCCAATCCCAAGACCTGAAATTAATA 2177
1328 LeuGluThrAspGlyThrSerMetMetLysValLeuSerGluArgAspVa 1344
2178 GGATGATCTTCAACACAAAGAGTGTTCAGCAGAACATCTCTCGGGAG 2227
1344 lAspProIle.....ArgThrSerSerAsn.. 1352
2228 GCTCATTTGTGCTTCTGATGTCCTCCAAATGCTCCCATACCTGATCTCTC 2277
1353AspIleCysGluIlePhe 1358
2278 CCACCTGCCCAATCACCGACCTGAAGCG..... 2307
1359GlnValLeuGlyIleGluAlaValArgLysSerValGluLy 1372
2308GAAATTCACGGG..... 2319
1372 sGluMetAsnAlaValLeuGlnPheTyrGlyLeuTyrValAsnTyrArgH 1389
2320GGCAGTCTCATTAAT 2334
1389 lSLeuAlaLeuLeuCysAspValMetThrAlaLysGlyHisLeuMeta 1405
2335 CTGACTTGGACAGCTCTCGGGATGATGATGACCATGGAACAGCTCACA 2384
1405 lIleThr.....ArgHisGly..... 1410
2385 GPATATCATTCGAATAAGTACAAGTATCTGATCTCAGACAGAGTTCA 2434
1411lIleAsnArgGlnAspThrGlyAlaLeu...MetArgCysSerPheG 1425
2435 ATGAATCTCTTCAAGTGAATACTACTGCTCTCATC..... 2469
1425 luGluThrValAspValLeuMetAspAlaAlaAlaHisAlaGluThrAsp 1441
2470 CCAAAGGAAGCAACTCTCAGGAAGTCTTTTGTAAACACCAAAACAT 2519
1442 ProMetArgGlyValSerGluAsnIleIleMetGlyGlnLeuProLysMe 1458
2520 TACTTTTGAATGGCAGAGATCTTTTCATTCCTGCTATTTCAGGCTGTGATA 2569
1458 t.....GlyThrGlyCysPheAspLeuLeuAspAlaGluL 1471
2570 AGGTGCTGCTGAAATCAGAAATATCAAC.....ATT 2601
1471 yScysArgPheGlyIleGluIleProAsnThrLeuGlyAsnSerMetLeu 1487
2602 GCACGAGTATCTTTGTTATT.....CCTCCACAGACTCC 2636
1488 GlyGlyAlaAlaMetPheIleGlyGlySerThrProSerMetThrPr 1504
2637 GCCAGAGACACCT.....AGTCTGATGAACGCTCTGCTC 2671

KW Cell adhesion; Signal; Glycoprotein; Transmembrane; Calcium-binding;
KW Repeat; EGF-like domain.

Length: 1195

3156 aGlyLeuAsnArgLysIleLeuTyrSerLeuIleAspSerAlaAspGlyG 3177

1460 ATGAGCGTGTCTCTCAGCGCTCC.....ATCCAGCTTGAG.....1494

3173 InPheSerIleasnGluLeuSerGlyIleIleGlnLeuGluLysProLeu 3189

14951511

3190 AspArgGluLeuGlnAlaValTyrThrLeuSerLeuLysAlaValAspGI 3206

1500 GGGATTACCCTCCAGAACACCCAGTGGATGCAATGGCACAGTGCCTGG 1549

3206 ngIleLeuProArgLeuThr.....AlaThrGlyThrValIleValS 3221

1550 AC.....1551

3221 erValLeuAspIleAsnAspAsnProProValPheGluTyrArgGluTyr 3237

1552 ...ACACCGTGGGAAAGGACACACTTTGTTCTATCACCTGGACACGCA 1598

3238 GlyAlaThrValSerGluAspIleLeuVal.....GlyThrGI 3250

1599 GCCTCCCAAAATCTCTCTCGGGATCCC.....1626

3250 uValLeuGlnValTyrAlaAlaSerArgAspIleGluAlaAsnAlaGluI 3267

1627AGTCGACAGAAGCAAGGTGGCTTTAGTGTGAC 1659

3267 leThrTyrSerIleIleSerGlyAsnGluHisGlyLysPheSerIleAsp 3283

1660 AAAACACCAAAATGGCC.....1677

3284 SerLysThrGlyAlaValPheIleIleGluAsnLeuAspTyrGluSerSe 3300

1678TACCTCCAAATC.....CCAGGCA 1696

3300 rHisGluTyrTyrLeuThrValGluAlaThrAspGlyGlyThrProSerL 3317

1697 TTGCTAAAGTGTGGCACTTGGAAATACAGTCTG.....1728

3317 euSerAspValAlaThrValAsnValAlaThrAspIleAsnAspAsn 3333

1729 ...CAAGCAGCTCACAAAACCTTGACCGCTGACTGTACGTCCCGTGGTCT 1775

3334 ThrProValPheSerGlnAspThrTyrThrThrValIleSerGluAspAl 3350

1776 CAATGCTACCTCGCTCCCAATACAGTACACTTCCAAACCAACAAGACA 1825

3350 aValLeuGluGlnSerValIleThrValMetAla.....AspA 3363

1826 CACGAAATATCCCAAGCCCTCGTGTAGTTTATGCAAAATATT.....1866

3363 spAlaAspGlyProSerAsnSerHisIleHisTyrSerIleIleAspGly 3379

1867 CGCCCAAGGAGCCTCC.....CCAATTC...AGGCCCAAGTGT 1901

3380 AsnGlnGlySerSerPheThrIleAspProValArgGlyGluValLysVa 3396

1902 CACAGCCCTGATT.....GAATCAGTCAATGGAAAAACAGTTTACCTTGG 1945

3396 lThrLysLeuLeuAspArgGluThrIleSerGlyTyrThrLeuThrValG 3413

1946 AACTACTGGATAATGCAGCAGGT.....1968

3413 InAlaSerAspAsnGlySerProProArgValAsnThrThrThrValAsn 3429

1969GCTGATGCTACTAAGGATGACGGTGTCTACTCAAGGATTTT 2009

3430 IleAspValSerAspValAsnAspAsnAlaProValPheSerArg.....3444

2010 CACAACCTTATCACACCAATGTTAGATACAGTGTAAAAGTCGGGGCTCTGG 2059

3445GlyAsnTyrSerValIleIleGlnGluAsnL 3455

[illegible]

```

FT DOMAIN          78 147 IG-LIKE C2-TYPE DOMAIN 1.
FT CARBOHYD      177 239 IG-LIKE C2-TYPE DOMAIN 2.
FT COXIN         274 338 IG-LIKE C2-TYPE DOMAIN 3.
FT DRAIN        366 428 IG-LIKE C2-TYPE DOMAIN 4.
FT FOMAIN       467 564 FIBRONECTIN TYPE-III 1.
FT FOMAIN       567 660 FIBRONECTIN TYPE-III 2.
FT FOMAIN       661 760 FIBRONECTIN TYPE-III 3.
FT FOMAIN       766 860 FIBRONECTIN TYPE-III 4.
FT FOMAIN       881 981 FIBRONECTIN TYPE-III 5.
FT DCMAIN      982 1083 FIBRONECTIN TYPE-III 6.
FT DCMAIN     1149 1153 POLY-VAL..
FT CARBOHYD    184 84 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD   221 221 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD   337 337 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD   501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD   520 520 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD   670 670 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD   746 746 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD   940 940 MISSING (IN ISOFORM 2).
FT VARSPLIC   442 461 MISSING (IN ISOFORM 3).
FT VARSPLIC   863 878 MISSING (IN ISOFORM 4).
FT VARSPLIC  1086 1096 MISSING (IN ISOFORM 5).
FT VARSPLIC  1279 1331 MISSING (IN ISOFORM 5).
SQ SEQUENCE    1493 AA; 163159 MW; 441DE919D5E17C0E CRC64;

alignment_scores:
Quality: 129.00 Length: 742
Ratio: 0.430 Gaps: 42
Percent Similarity: 40.431 Percent Identity: 20.620

alignment_block:
US-09-049-696-18 x NEOL_MOUSE ..

Align seg 1/1 to: NEOL_MOUSE from: 1 to: 1493

898 CCACCAATCCACCTTCTCATTCGATCGAGTTGGACAAAAGAATTGTGTG 947
||||| ||||| :::: | ||||| :::: : |||||
21 PPTPTPTPTProLeuLeuLeuLeuProLeu.....LeuLeuLe 34

948 TTTATCTCTTGACAATACTGGAACATCGGCAGCHGTGAACGCCCTCAATC 997
||| ||||| :::: | ||||| :::: : |||||
34 uLeuGlyArgProAlaSerGlyIAlaAlaThrLysSerGlyProArgA 51

998 GACTGAATCAGACGCCAGCTT .....TTCCCTG 1026
||| ||||| :::: | |||||
51 rGlnSerGlnGlyAlaSerValArgThrPheThrPheThrPheLeu 67

1027 CTGCAGACAGTTGAGCTGGGCTCTGGGTTGGGATGTGACATTTGACAG 1076
::::: |||
63 ValGluProVal..... 71

1077 TGCTCCCCATGTACAACTGAATCATACATCAAACAGTGGCAGTCACA 1126
71 ..... 71

1127 GGGACACACTGCCAAAAAATTACTCTGCAGCAGCTTCAGCAGGAGCATCC 1176
||||| ||||| :::: | ||||| :::: : |||||
72 AspThrLeuSerValarg .....GlySerSer 80

1177 ATC.....TGACGGGGCTGCATCGGCATTTACTGTGATTAGGAA 1217
||||| :::: | |||||
81 ValIleLeuAsnCySerAlaTyrsrGluProSerProAsnIle..... 95

1218 GAATATCCAATGATGATCTGAATTTGTCTGCTGACGATGGGGAAG 1267
::::: ||||| :::: | ||||| :::: : |||||
96 GlutriPylsLyaspGlyThrPheLeuAsnLeuGluSerAspArg. 111

1268 ACAACACTATAAGTCGGTCTTTAACAGGAGCTCAACAAAGTGTGGGCC... 1314
::::: ||||| :::: | |||||
112 .....ArgGlnLeuLeuProAspGlySerLeu 120

1315 .....ATCATCCACACAGTCGTTTTGGGGCCCTCTGCAGGTCA 1352

```

121 PheileSerAsnValValHisSerLysHisAsnLysProAsp..... 134
1353 AGAACTAGAGGAGCTGTCCTCAAAATGACAGGAGTTTACAGACATATGCTT 1402
135GluGlyPheTyrGlnCysValAlaT 143
1403 CAGATCAAGTTTCAACAATAGGCGCTCATGATGCTTTTGGGCGCTTCA 1452
143 hr.....ValAspAsnLeuGlyThrIleVal 151
1453 TCAGGAAATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGG 1502
152 Ser.....ArgThrAlaLysLeuThrValAlaGl 161
1503 ATTAACCTCCAGAACAGCCAG.....TGGATGAATG 1534
161 yLeuProArgPheThrSerGlnProGluProSerSerValTyrValGlyA 178
1535 GCACAGTCATCGTCGACAGACCGTGGGAAGGACACTTGTCTTTCATC 1584
178 snSerAlaIleLeuAsnCysGluValAsnAlaAspLeuValProPheVal 194
1585 ACCGTCGACAGCGCTCCCAATCTCTCTGGGAT..... 1623
195 ArgTrrpGluGlnAsnArgGlnProLeuLeuAspArgIleValLy 211
1624CCAGTGGGA.....CAGAAGCAAGGTG 1645
211 sLeuProSerGlyThrLeuValIleSerAsnAlaThrGluGlyAspGlyG 228
1646 GCTTT.....GTAGTGGACAAAACACCAAAATGGCGCTACTCCAA 1686
228 lyLeuTyrArgCysIleValGluSerGly..... 237
1687 ATCCAGGCACTGTGAAGTTGGCACTTGGAAATACAGT..... 1725
238GlyProProLysPheSerAspGluAlaGl 247
1726CTGCAGCAGCTCACAACTTGACCTGACTGCTCAGT 1765
247 uLeuLysValLeuGlnAspProGluGluIleValAspLeuValPheLeuM 264
1766 CCCGTGGCTCC.....AATGCTACCTCGCT 1791
264 etArgProSerSerMetLysValThrGlyGlnSerAlaValLeuPro 280
1792 CCAATTACAGTACTTCCAAAACGAACAGGACACCAAGCAAAATTCGCCAG 1841
281 CysVal.....ValSerGlyLeuProAl 288
1842 CCCTCTGGTAGTTATGCAATATTCGCAAGGAGCTCCCAATTCCTCA 1891
288 aProValValArgTrrpMetLysAsnGlu..... 298
1892 GGGCCAGTGTACAGCCCTGATTGAATCAGTGAATGGAACAACTTACC 1941
299ValLeuAspThrGluSerSerGlyArgLeuValLeu 310
1942 TTGGAACCTACTGGATAATGGAGCAGGT.....GCTGATGC 1976
311 Leu.....AlaGlyCysLeuGluIleSerAspVa 321
1977 TACTAAGATGACGGTGTCTACTCAAGTATTCACAACCTTATACACAGA 2026
321 lThrGluAspAspAla.....GlyThrTyrPheCysIleAlaAsp..A 335
2027 ATGCTAGATACAGTCTAAAGTGGCGCTCTGGGAGGAGTTAAGCAGCC 2076
335 snGlyAsnLysThrValGluAlaGlnAlaGluLeuThrVal..... 348
2077 AGACGGAGAGTGATACCC.....CAGCAGAGTGGAGCAGCTGTAC... 2115

349GlnValProGlyPheLeuLysGlnProAlaAsnIleTyrAl 363
2116ATACCTGGC..... 2124
363 aHisGluSerMetAspIleValPheGluCysGluValThrGlyLysProT 380
2125TGGATTGAGAATGATGAATACAAATGAATCCACCA 2160
380 hrProThrValLysTrrpValLysAsnGlyAspValValIleProSerAsp 396
2161 AGACCTGAAATTAATAAGGATGATTTCAACACAAAGCAAGTGTGTTTCAG 2210
397 AsnPhelLysIleValLysGlu.....HisAsnLeuGlnValLeuGlyLe 411
2211 CAGAACATCTCTGGGAGGCTCATTT.....GTGGCTTCTGATGTC 2251
411 uValLysSerAspGluGlyPheTyrGlnCysIleAlaGluAsnAspValG 428
2252 CAATGCT.....CCC 2262
428 lyAsnAlaGlnAlaGlyAlaGlnLeuIleLeuGluHisAspValAla 444
2263 ATACCTGATCTCTTCCACCTGGCCAAATCACCC..... 2295
445 IleProThrLeu...ProProThrSerLeuThrSerAlaThrThrAspHi 460
2296GACCTGA 2302
460 sLeuAlaProAlaThrThrGlyProLeuProSerAlaProArgAspValV 477
2303 AGCGGAAATTCACGGGGCAGTCTCATTAATCTCAGCTTGGACAGCTCCT 2352
477 alAlaSerLeuValSerThrArgPheIleLysLeuThrTrrpArgThrPro 493
2353 GGGGATGATTATGACCATGGAAACAGCTCACAAAGTATATATTCGAATAAG 2402
494 AlaserAsp...ProHisGlyAspAsnLeuThrTyrSerValPheTyrTh 509
2403 TACAAGTATCTTGTATCTCAGACAGCAAGTTCAATGAATCT..... 2442
509 rLysGluGlyValAspArgGluArgValGluAsnThrSerGlnProGlyG 526
2443CTTCAAGTGAATACTACTGCTCTCATCCCAAGGAAGCAACTCTGAG 2490
526 luMetGlnValThrIleGlnAsnLeuMetPro.....Ala 537
2491 GAAGTCTTTTGTGTTAAACCAAGAAACATTACTTTTGAATAATGGCACAGA 2540
538 ThrValTyrIlePheLys..... 543
2541 TCTTTTCTATTGCTATTTCAGGCTGTTGATAAGTCTGATCTGAATCAGAAA 2590
544ValMetAlaGlnAsnLysHisGly...SerGlyGluS 555
2591 TATCCAACATTCGACAGATATCTTTTGTGTTTATTCTCCACACAGACTCCGCCA 2640
555 erSerAlaProLeuArgVal.....GluThrGlnPro 565
2641 GAGACACTAGTCTGATGAAACGCTGCTCTCTGCTCTCTATATTCATAT 2690
566 GluValGlnLeuPro.....GlyProAlaProAsnIleArgAl 578
2691 CAACAGACCACTTCTCGCATTCACATTTTAAATAATATGTGGAG... 2736
578 atyAlaThrSerPro.....ThrSerIleThrValThrTrrpGluThrP 593
2737TGGATAGGAGAACTGCAG 2754
593 roLeuSerGlyAsnGlyGluIleGln 601

This Page Blank (uspto)

seq_name: pir2:C81675

seq_documentation_block:
 A:polymorphic membrane protein B/C family TC0694 [imported] - Chlamydia muridarum, Chlamydia trachomatis MoPn
 C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 1
 C:Accession: C81675
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.;
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.;
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia muridarum
 A:Reference number: A81500; MUID:20150255
 A:Accession: C81675
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1672 <TET>
 A:Cross-references: GB:AE002338; GB:AE002160; NID:g7190724; PIDN:
 A:Experimental source: strain Nigg (MoPn)
 C:Genetics:
 A:Gene: TC0694

alignment_scores:
 Quality: 153.50 Length: 802
 Ratio: 0.468 Gaps: 34
 Percent Similarity: 40.898 Percent Identity: 18.080

alignment_block:
 US-09-049-696-18 x C81675 ..

Align seg 1/1 to: C81675 from: 1 to: 1672

```

547 TTCTACTTTCCTCAATGGAAGATCAAGCAGTAGTAAAGATGTTTCAGCAGGTAT 596
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
122 PheTyAsnGluHisSerGlyIleLeuSerPheMetAlaArgSerGlyVa 138
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
597 TACTGGT.....ACAATGTAGTAAGAAGTGTCTCAGGAGGCA 634
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
138 lGluGlySerLeuThrLeuSerAsnIleLysMetThrGlyAspGlyVa 155
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
635 GCTGTATACACAAA...AGATGCACATTCATAAAGTAACAGCAGCTCTAT 681
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
155 lalleTySerGlnGlyGluLeuPheThrAspLeuThrGlyLeuThr 171
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
682 GAAAAAGAGTGTAGTTGTTCTCCAAATCCCGCCAGCAGCAGGAGGCTTC 731
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
172 lIeGlnGly.....AsnLeuSerGlnLeuSerGlyGlyG1 183
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
732 TATAATCTTTGCACACATCTGTGATCTCTATAGTTGAATCTGTACAGAAC 781
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
183 yIlePheGlyGly.....SerThrIleSerPhe..... 192
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
782 AAAACCCACAAAGAGAGCTCCAAACAGCAAAATCAAAATGCAATCTC 831
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
193 .....SerglyIleAsnGlnAlaThrPheSer 201
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
832 CGAAGCACATGGGAAGTATCCGCTGATCTCAGGACTTTAAGAAACCAC 881
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
202 SerAsnThrAlaGluValValProGlu.....GluThr 213
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
882 TCCT.....ATGACAAACACAGCCACCACCAATCCCA 910
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
213 rProAsnProAsnProGlyThrGlnThrThrThrSerGlnProSerProt 230
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
911 CCTCTCATTTGCTGCAGATGGACAAAGAAATTTGTGTTTAGTCCTTCAG 960
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
230 hr.....SerLysValGlnSerLeuPheThrTyr 239
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
961 AAATCTGGAAGCAGTCGCTGTAACCGCTCAATCGACTGAATCAAGC 1010
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
240, SerSerSerThrGlnAlaAsnGlyAsn..... 248

```

1011 AGCCAGCTTTCTCTGTCAGACAGTGTGAGCTGGGCTCCTGGTTGGGA 1060
 248
 1061 TGGTGACATTTGACAGTGTGCTGCCCATGTACAAAAGTGAATCATACAGATA 1110
 248
 1111 AACAGTGGCAGTGACAGGGGACACACTGCGCCAAAGATTACTGCAGCAGC 1160
 249GlyAlaAspSerGlnThrProSerHisLys.....ProG1 260
 1161 TTCAGGAGGACCTCCATCTGCAGCGGCTTCGATCGCATTTACTGTGA 1210
 260 ySerGlyAlaIleThrGly.....AspLeuThrIle. 273
 1211 TTAGGAAGAAATATCCAACTGATGATCTGAAATTTGTGCTGCACGGAT 1260
 274SerAspSerGlnGluIleValPhe..... 281
 1261 GGGAGACAACTATAAGTGGTGTCTTTAACGAGTCAAAAGTGG 1310
 282SerValAsnLysAlaSerLysAspG1 290
 1311 TGCCATCATCCACACATCCTCTTGGGCCCTCTGCAGCTCAAGAACTAG 1360
 290 yGlyAlaIle..... 293
 1361 AGGAGCTGTCCAAATGACAGAGGTTTACAGACATATGCTTCAGATCAA 1410
 294PheAlaGluLysAsnValSerPheGluAsnIleThrThrLeuLys 308
 1411 GTTCAGAACAAATGCCCTCATTTGATGCTTTTGGGCCCTTTTCATCAGGAAA 1460
 309 ValGlnAsnAsn.....GlyAlaGluLysGlyG1 319
 1461 TGGAGCTGTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTACCC 1510
 319 yGlyIleThrAlaSerGlyAspLeuSerIleGlnSerSerLysGlnSerL 336
 1511 TCCAGAACACCCAGTGGATGAATGGCAGTGTGCTGGGACAGCACCCTG 1560
 336 euPheAsnSer..... 339
 1561 GGAAGACACTTTGTTTCTTATACCTGGACACAGCAGCTCCCAAAAT 1610
 339 339
 1611 CCTCTCTGGATCCCAAGTGGACAGAGCAAGGTGGCTTTGTAGTGACA 1660
 340AsnThrSerLysGlnGlyGlyAlaLeuThrIleGluG 353
 1661 AAAACACAAATGGCTACCTCCAAATCCAGGCATTTGTAAGTTGGC 1710
 353 lyAsnValAspPheLysAspLeuGluIleArgIle..... 365
 1711 ACTTGGAAATACACTGCAAGCAAGCTCAAAACCTTGACCTGACTGT 1760
 366LysTyrAsnLysSerGlyThrPheGluThrLysLysValThrLe 380
 1761 CAGTCCCGTGCCTCAATCCTACCCTGCCCTCCCAATTTACAGTCACTCCA 1810
 380 uSerLeuProGluAlaGlnThrAsnLysSerSerValThrAlaAlaSerG 397
 1811 AAACGAACAAGGACACAGCAAAATTCGCCAGCCCTCTGGFA..... 1851
 397 InSerGlyProAsnThrThrProThrProThrProValThrAlaLys 413
 1852GTTTATGCAATATTCGCCAAGGACCTCCCAATTTCTCAG 1892
 414 GlyGlyLeuThrThrGluLysAsnLeuSerIleSer..... 426
 1893 GGCCAGTGTACAGCCCTGATTGAATCAGTGAATGGNAACAGTTACCT 1942

427AsnIleThrGlyIleIleGluIleThrAsnAsnLysAlaThr.... 440
 1943 TGGAACTACTGTAATGAGCAGAGGTGCT..... 1971
 441AspValGlyGlyAlaThrValLysGlyThrLeuThr 453
 1972GATGCTACTAA 1982
 454 CysLysAspSerHisArgLeuGlnPheGlnLysAsnSerSerGluLysLy 470
 1983 GGATGAGGCTGTACTCAAGGTATTTCAACAATTTATGACACGAATGGTA 2032
 470 sGlyGlyLeuThr.....ThrGluAspThr..... 479
 2033 GATACAGTGTAAAGTGGCGCTCTGGAGGAGTTAACCGACGACGACGG 2082
 480IleThrLeuSerAsnLeuThrGlyLysThrLeuPheGlnGlu 493
 2083 AGAGTGATACCCAGCAGAGTGGAGCACTGTACATACCTGGCTGGATTGA 2132
 494 AsnThrAlaLysGluGlyGlyGlyLeuThrIleGlnGly..... 507
 2133 GAATGATGAATACATGGAATCCACCAAGACCTGAAATTAATAAGGATG 2182
 508AspA 509
 2183 AT.....GTTCAACACAAGCAAGTGTGTTTC..... 2208
 509 spLysThrLeuThrMetThrGlyLeuAspSerPheCysLeuIleAspAsn 525
 2209 ...AGCAGAACATCCTCGGAGGCTCATTTGTGGCTTCTGATGTCCTCAA 2255
 526 ThrSerAlaThrHisGlyGlyGlyAlaThrValThrLysGluIleSerG1 542
 2256 TGCTCCCATACCTGATCTC.....TTCACACCTGGCCAAATCACCCACC 2299
 542 nThrTyrThrSerAspValGluGluPhePro.....GlyIleThr.... 555
 2300 TGAAGCGGAAATTCACGGGGGAGTCTCATTAATCTGACTTGGACAGCT 2349
 556ProValHisGlyGluThrIleIleSerGlyAsnLysAlaThr 569
 2350 CTGCGGATGATTATGACCATGGAACAGCTCACAAATATATCATTCGAAT 2399
 570 GlyLysSerGlyGlyValCysThrLysHisLeuValLeuSerAsnLe 586
 2400 AAGTACAAGTATTCTGATCTCAGAGACAGTTCATGAATCTCTTCAAG 2449
 586 uGlnThrIleSerIleSer..... 592
 2450 TGAATACTGCTCTCATCCCAAGGAAGCAACTCTGAGGAAGTCTTT 2499
 592 592
 2500 TTGTTTAAACCAGAAACATTAATCTTTTGAATATGGCACAGATCTTTTCAT 2549
 593GluAsnPheAlaSerGluAsnGlyGly..... 602
 2550 TGCTATTACAGCTGTTCAATGAGTCTGATCTGAAATCAGAATATATCCAACA 2599
 603 AlaCysThrCysProAspAsnPhe..... 610
 2600 TTGCACAGTATCTTTGTTTATTCCTCCACAGACTCCGCCAGACACACCT 2649
 611ProAlaProThrAlaSerThrPro 618
 2650 AGTCCTGATGAACGCTGCTGCTCT.....TGTCCTAA 2681
 619 SerThrAsnGlnThrAlaAlaProLysAspLysAspPheLeuIleAs 635
 2682 TATTATATCAACAGCACCATT.....C 2704

326 ATCTGATGTTCTGCTGAGTCTACTCTCCAGTAAATGATGAACCC 375
|| : : : : : ||| : : : : :
216 sn.....LeuPheSerGlnHisThrLeuHisArgThrAspGlu... 229
376 TACACTGACGATGGCAACTGTGAGAGACAGGGTCAAGAGTCCACCT 425
||| : : : : : ||| : : : : :
230 ...ArgThrGlnHisAsn.....GlyLeuGlyTrpArg...HisPh 242
426 CACTCTGATTTTCATTCAGGAAAAAGTAGCTGAATATGGACCACAAG 475
:
242 eThrProThrMetSerGlyIleAsnPh..... 252
476 GTAGGCCATTTGTCATGAGTGGGCT.....CATCTAGATGGGAGTA 519
||| : : : : : ||| : : : : :
253PhePheHisAspLeuSerArgTrpHisSerArgAlaGlyIle 267
520 TTTCAGCAGTACATAATGATGAGAAATCTACTTATCAATGGA..... 564
||||| : : : : : ||| : : : : :
268 GlyAlaGlyTrpArgAspTrpLeuLysLeuSerSerAsnGlyTrpLe 294
565 .AGAATACAGCAGTAGTAAATGTTTCAGCAGTATTACTGGTACAAATGTAG 613
||| : : : : : ||| : : : : :
284 uArgLeuThrAsnTrpArgSerAlaProGluLeuAspAsnAspTyrGluA 301
614 TAAAGAGTGTCAAGGA..... 630
301 leArgProAlaAsnGlyTrpAspValArgAlaGluGlyTrpLeuProAla 317
631GGCAGCTTTACACCAAAAGATGCACATTCATAAAGT 668
318 TrpProHisLeuGlyLysLeuValTyrGluGlnTyrTrpGlyAspG 334
669 AACAGGACTCTATGAAAAGAGTGTGAGTTTCTTCCAATCCGCCAGA 718
:
334 uValAlaLeuPheAspLys.....AspAspArgGlnS 345
719 CGGAGAGGCTTCTATATGTTTGCACACATGTTGAT.....TCTATA 762
:
345 eAsnProHisAlaIleThrAlaGlyLeuAsnTyrThrProPheProLeu 361
763 GTTGAATTTCTGACAGACAAAACCAACAAAGAGTCCCAACCAAGCA 812
||| : : : : : ||| : : : : :
362 MetThrPheSerAlaGluGlnArgGlnGlyLysGlnGlyGluAsn..... 376
813 AAATCAAAATGCAATCTCCGAGCAGCATGGAA..... 846
:
377 .AspThrArgPheAlaValAspPheThrTrpGlnProGlySerAlaMetG 393
846 846
393 lnLysGlnLeuAspProAsnGlnuValAspAlaArgSerLeuAlaGly 409
847GTGATCCGTGATTTCTGAG.....GACTTTAA 872
||| : : : : : ||| : : : : :
410 SerArgPheAspLeuValAspArgAsnAsnIleValLeuGlyTrpArg 426
873 GAAA.....ACCACCTCTATGACA.....A 892
:
426 glyLysGluLeuValArgLeuThrLeuThrAspProValThrGlyLys 443
893 CACAGCCCAAAATCCACCTTCTCATTTCTGCTGAGATT..... 930
:
443 eArgGlyValLysSerLeuValSerSerLeuGlnThrLysTyrAlaLeu 459
931 ...GGACAAGAAATTCGTGTTAGTCCCTTGACAAAATCTGGAAGC..... 972
||| : : : : : ||| : : : : :
460 LysGlyTyrAsnValGluAlaThrAlaLeuGluAlaAlaGlyLysVa 476
973 .ATGGCGACTGGTAACCGGCTC..... 993
:
476 lValThrThrGlyLysAspIleLeuValThrLeuProAlaTyrArgPhe 493

993 993
493 hrSerThrProGluThrAspAsnThrTrpProIleGluValThrAlaGlu 509
994AATCGACTGATCAAGCAGGCGACGCTTTCTCTGCTGCA 1031
||| : : : : : ||| : : : : :
510 AspValLysGlyAsnPheSerAsnArgGluInSerMetValValGlu 526
1032 GACAGTTGAGTGGGCTCTGGTGGGATGGTGCATTTGACAGTTCAGACGTGCTG 1081
:
526 nAlaProThrLeuSerGln.....LysAspSerSerV 537
1082 CCCATGTACAAGTGAACATACACATAAACAGTGCAGCTGACAGGGAC 1131
:
537 alSerLeuSerSerGlnThrLeuSerAlaAspSerHisSer..... 550
1132 ACATCGCCAAAAGATTACTCGCAGCAGCTTCAGGAGGAGGCTCCATCTG 1181
||| : : : : : ||| : : : : :
551 ThrAlaThrLeuThrPheIleAlaHisAspAlaAlaGlyAsnProValII 567
1182 CAGCGGGCTTCGATCGCATTTACTGTGATTAGGAAGAAATATCCAACCTG 1231
||| : : : : : ||| : : : : :
567 e...GlyLeu.....ValLeuSerThrArgHis.....G 576
1232 ATGATCTCAAAATTGTCTGCTGACGATGGGGAAGACACACTATAAGT 1281
:
576 lGlyValGlnAspIleThrLeuSerAspTrpLysAspAsnGly...Asp 591
1282 GGTGTCTTTAACGAGGTCAAAAGTGTGGCCTCATCCACACAGTCGC 1331
||| : : : : : ||| : : : : :
592 GlySerTyrThrGlnIleLeuThrThrGlyAlaMetSerGlyThrLeuTh 608
1332 TTGGGGGCC.....TCTGCAGCTCAAGAAGTCTAG 1360
||| : : : : : ||| : : : : :
608 rLeuMetProGlnLeuAsnGlyValAspAlaAlaLysAlaProAlaVal 625
1361 AGAGAGTGTCCAAAATCAGAGGAGTTTACAGACATATGCTTCAGATCAA 1410
:
625 alAsnIleLeuSerValSerSer...ArgThrHisSerIleLys 640
1411 GTTCAGAACAAATGGCTCATTTGCTTTTGGGGCCCTTTTCATCAGGAAA 1460
:
641 IleAspLysAspArgTyrLeu.....SerGlyAs 650
1462 TGGAGTGTCTCAGCGTCCATCCAGCTTGAGATTAAGGATTAAACCC 1510
:
650 nProIleGluValThrValGluLeuArgAspGluAsnAspLysProVal 667
1511 TCCAGAACACCCAGTGGATCAATGGCAGCAGTGTGCTGGAC..... 1551
:
667 ysgluGlnLysGlnGlnLeuAsnThrAlaValSerIleAspAsnValLys 683
1552AGCACCGTGGGAAAGGACACT.....TTGTTTCTTAT 1583
||| : : : : : ||| : : : : :
684 ProGlyValThrThrAspTrpLysGluThrAlaAspGlyValTyrLysAl 700
1584 CACTGGACAACGACG.....CCTCCCCAAATCCTTC 1615
||| : : : : : ||| : : : : :
700 aThrTyrThrAlaTyrThrLysGlySerGlyLeuThrAlaLysLeuLeuM 717
1616 TC.....TGGGATCCAGTGGAGCAACAGCAGTGGCTTTGTAGTGGAC 1659
:
717 etGlnAsnTrpAsn...GluAspLeuHisThrAlaGlyPheIleLeuAsp 732
1660 AAAACACCAAAATGGCTTACCTCCAAATCCAGGCAATTCAGGTGG 1709
||| : : : : : ||| : : : : :
733 AlaAsn.....ProGlnSerAlaLysIleAl 741
1710 CACTTGGAAATACAGT.....CTGCAAGCAAGCTCAC 1741
||| : : : : : ||| : : : : :
741 aThrLeuSerAlaSerAsnAsnGlyValLeuAlaAsnGluAsnAlaAla 758
1742 AAACCTTGACCTGACTGTACGCTCCGCTCGCTCCCAATGCTACCTGCCT 1791

```

:::|||||:::|||||:::|||||:::|||||
758 snThrValSerValAsnValAlaAspGluGlySerAsn..... 770
1792 COAATT.....ACAGTGAC 1805
|||||
771 ProIleAsnAspHisThrValThrPheAlaValLeuSerGlySerAlaTh 787
1806 TTCAAACGACACAGGACACACAGAAA..... 1833
|||||
787 rSerPheAsnAsnGlnAsnThrAlaAlaLysThrAspValAsnGlyLeuAlar 804
1833 ..... 1833
804 hrPheAspLeuLysSerSerLysGlnGluAspAsnThrValGluValThr 820
1834 .....TTCGCCAGCCCTCTGGTAGTT..... 1854
821 LeuGluAsnGlyValLysGlnThrLeuIleValSerPheValGlyAspSe 837
1855 .....TATGCAAAATTCGCAAGGAGCCCTCCCAATTCTC.... 1890
837 rSerThrAlaGlnValAspLeuGlnLysSerLysAsnGluValValAlaA 854
1891 .....AGGCCAGTGTACAGCCCTGATTGAATCACTGAATGGA 1929
|||||
854 spGlyAsnAspSerAlaThrMetThrAlaThrValArgAspAlaLysGly 870
1930 AAA.....ACAGTTACTTGGAACTACTGTGATAATGGAGC 1964
871 AsnLeuLeuAsnAspValLysValThrPheAsnVal.....AsnSerAl 885
1965 AGGTGCT.....CATGCTACTAAGGATGACGGTGTCTACT 1999
885 aAlaAlaLysLeuSerGlnThrGluValAsnSerHisAspGlyIleAlar 902
2000 CAAGGTTATTCACAACATTATGACACCAATGCTAGATACACTGATAAGTG 2049
902 hrAlaThrLeuThrSerLeu...LysAsnGlyAspTyrThrValThrAla 917
2050 CGGGCTCTGGAGGAGTAAACGACGACGACGAGAGTGATA..... 2091
918 SerValSerSerGly...SerGlnAlaAsnGlnGlnValIlePheIleG 933
2092 .CCCCACAGAGTGAGGACACTGTACATACCTGGCTGGATTGACATGATG 2140
|||||
933 yAspGlnSerThrAlaAlaLeuThrLeuSer.....ValProSerGlyA 948
2141 AAATACAATGGAATCCACCAAGACCT..... 2166
948 spIleThrValThrAsnThrAlaProLeuHisMetThrAlaThrLeuGln 964
2167 GAAATTAATAAGGATGATGTTCAACACACAGCAAGTGTGTTTCAGC..... 2211
965 AspLysAsnGlyAsnProLeuLysAspLysGluIleThrPheSerValPr 981
2212 .....AGAA 2215
981 oAsnAspValAlaSerArgPheSerIleSerAsnSerGlyLysGlyMetT 998
2216 CATCTCGGAGGCTCATTTCTGGCTTCT..... 2244
998 hrAspSerAsnGlyThrAlaIleAlaSerLeuThrGlyThrLeuAlaGly 1014
2245 .....GATGTCGCAATGCTCCCATCTGATCTCTT 2276
1015 ThrHisMetIleThrAlaArgLeuAlaAsnSerAsnValSerAspThrG 1031
2277 CCCACCTGGCCAAATCACCAGC..... 2298
1031 nProMetThrPheValAlaLysAspArgAlaValValLeuGlnT 1048
2299 ..CTGAAGCGGAAATTCACGGGGCGAGTCTCTAATCTGACTTGGACA 2346
|||||

```

```

1048 hrSerLysAlaGluIleIleGlyAsnGlyValAspGluThrThrLeuThr 1064
2347 GTCCTCGGGGATGATTATGACCATGGACAGCTCACAACTATATCATTCG 2396
|||||
1065 AlaThrValLysAspProPheAspAsnValValLysAsnSerValVa 1081
2397 AATAAGTACAAGTATTCTTGATCTCAGACACAAAGTTCAATGAATCTCTTC 2446
|||||
1081 lPheArgThrSerProAlaAspThrGlnLeuSerLeuAsn...AlaArgA 1097
2447 AAGTGAATACTACTGCTCTCATCCCAAGGAAGCCAACTCTCGAGGAATC 2496
1097 snThrAsnGluAsnGlyIleAlaGluValThrLeuLysGlyThrValLeu 1113
2497 TTTTGTGTTTAAACCCAGAAAACATTACTTTTGAATAATGGCAGATCTTTT 2546
|||||
1114 GlyValHisThrAlaGluAlaIleLeuLeuAsnGlyAsnArgAspThrLy 1130
2547 CATTCGTATTTCAGGCTGTTGATAAGGTCGATCTGAAATCAGAAATATCCA 2596
1130 sile.....ValAsnIleAlaProAspAlaSerA 1140
2597 ACATTGCGACGAGTATCTTCTTTTATCTCCACAGACTCCGCCAGAGACA 2646
|||||
1140 sn...AlaGlnValThrLeuAsnIleProAlaGlnValValThrAsn 1155
2647 CCTAGT.....CCTGATGAAACGCTCTGC 2669
1156 AsnSerAspSerValGlnLeuThrAlaThrValLysAspProSerAsnH 1172
2670 TCCTGTGCTTAATATTTCATATCAACAGCAGCATTCCT 2706
1172 sproValAlaGlyIleThrValAsnPheThrMetPro 1184

```

seq_name: pir2:C81265

```

seq_documentation_block:
probable lipoprotein Cjl678 [imported] - Campylobacter jejuni (strain NCTC 11168)
C:Species: Campylobacter jejuni
C:date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C:Accession: C81265
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba
Nature 403, 665-668, 2000
A:title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: AB1250; MUID:20150912
A:Accession: C81265
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-928 <PAR>
A:Cross-references: GB:AL139079; GB:AL111168; NID:g9698971; PIDN:CAB73665.1; PID:g696
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cjl678

```

```

alignment_scores:
Quality: 151.50 Length: 913
Ratio: 0.384 Gaps: 42
Percent Similarity: 43.264 Percent Identity: 18.401

alignment_block:
US-09-049-696-18 x C81265 ..
Align seg 1/1 to: C81265 from: 1 to: 928

```

```

298 GTGACACCAAACTTCAGACCTACAAAATCCTGATGTTCTGGTTGCTGA 347
50 IleAsnThrLysIleAspThrPheThrAsnAsnGlyPheIleAsnSerPr 66
348 GTCTACTCTCCAGGTAATGATGAACCCCTACACTGAGCAGATGGCAACT 397
66 oGlySerGlyGlnTrpAsnAsnGlyIleTrp.IleSerSerAsnAlaThr 82

```

[illegible]

1069TTTGACAGCTCCTCCCATGTACAAAG 1094
319 AlaSerIleGlySerAspIleGlyLeuGlyAlaLysValGluG1 335
1095 TGAACATCATACAGATAA 1111
335 yGlyLeuValAsnGlnGlyAsnGlySerIleSerGlySerValGlnValS 352
1112 ACAGTGGCAGTACAGGACACACTCGCCAAAGATTACCTGCAGCAGCT 1161
352 erGlySerSerIleAspSerIleThrAsnGluGlyAsnGlyAlaIle 368
1162 TCAGAGGAGCAGCTCATCTGCAGCGGCTTCGATCGCATTTAC 1205
369 SerGlySerIleThrValThrLysAspSerLysLeuAspSerIleThrAs 385
1206 TGTGATTAGGAAGAAATATCCAACGTGATGGATCGAAATTGTGCTGCTGA 1255
385 nThrSerThrSerThrGlyIleSerGlySerIleI 398
1256 CGGATGGGGAAGACACACTATAAGTGGTGCTTTAAAGAGGTCAACAA 1305
398 hrAsnAsnSerAspAsnLysLeuGluIleSerAsn 409
1306 AGTGGTGCCATCATCCACACAGTCGTGTTGGGGCCCTCTCGAGCTCAAGA 1355
410 SerGlyAsnIle 413
1356 ACTAGAGGAGCTGTCCAAATGACAGGAGGTTACAGACATATG 1399
414 422
1400 CTTCAGATCAAGTTCAGAACATGCCCTCATGTGCTTTT 1440
422 erAlaAspMetValIleSerAsnSerAsnGlyGlyThrIleSerGly... 437
1441 GGGGGCCCTTCATCAGGAATAGGAGCTGCTCTCAGCGCTCCATCCAGCT 1490
438 GlyIleSerSerGlySerGlySerThrSerIleSerAsn 451
1491 TGAGAGTAAGGATTAACCTCCACAGACGACGAGTGAATGAGGCACAG 1540
452SerGlnGlySerThrIleAsnAsnGlyIleThrValSerGlySerA 467
1541 TGATCGTGACAGCACCGTGGAAAGGACACTTTGTTT 1578
467 laGlnValGluIleSerAsnGlnGlySerValGlyLysAsp 480
1579 CTTATCACCTGGACAAACGACGCTCCCAATCTCTCTGGGATCCCGAG 1628
480 480
1629 TGGACAGAAGCAGGTGGCTTTGTAGTGGACAAAACACCAAATGSCCT 1678
481GluAsnGlyAsnThrValThrAsnAsn 489
1679 ACCTCCAAATCCCAGGCATTCCTAAGGTGGCACTTGGAAATAC 1722
490GlySerGlySerValGlyIleLysAspTrpLeuVal 501
1723 ACTCTGCAAGCAGCTCAACAACCTTGACCTGACTGCTCAGTC 1766
502 SerThrAspLysAsnThrGlyLysLeuAsnThrValIleGlyGlySer 518
1767 CGGTGGCTCCATGCTACCTCGCTCCCATTAATACAGTACACTTCCAAAACGA 1816
518 rArgAlaPheAsnValLysValGluAsnIleThrValAspGlnSerA 534
1817 ACAAGGACACACAGCAAAATTCGCCAGCCCTCTGGTAGTTTATGCAAATATT 1866
534 snValAspLeuGluLeu 540
1867 CGCCAAAGGAGCTCCCAATTTCTCAGGCGCCAGTGTCTACAGCCCTGATTGA 1916

C:Accession: B85547
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 iller, D.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
 Nature 409: 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: B85547
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-5188 <STO>
 A:Cross-references: GB:AE005174; NID:g12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: Z0615

alignment_scores:
 Quality: 145.50 Length: 554
 Ratio: 0.560 Gaps: 28
 Percent_Similarity: 46.931 Percent Identity: 21.841
 alignment_block:
 US-09-049-696-18 x B85547 ..
 Align seg 1/1 to: B85547 from: 1 to: 5188
 964 TCTGGAAGCATGCGAGCTGGTAACCCCTCAATCGACTGAATCAACGAGG 1013
 3362 ThrGlyAlaValAla...GlyAspArgLeu...ValValThrIleAlaG1 3376
 1014 CCAGCTTTCTCCTCGCAGACAGTTGAGCTGGGGTCTCTGGTT..... 1056
 3376 yGlnGlnTyThrValThrSerThrAspAlaSerGlyAsnTrpSerValGlyV 3393
 1057GGGATGGTGACATTT... 1071
 3393 alProAlaSerValIleSerGlyLeuAlaAspGlyThrValThrIleSer 3409
 1072GACAGTGTGCGCCAT.....GTACAAAGTGAACATCAT 1103
 3410 AlaThrIleThrAspSerAlaGlyAsnSerSerThrGlnThrHisAsnVa 3426
 1104 ACAGATAAAGCAGTGGCAGT.....GACAGGG 1129
 3426 lGlnValAsnThrAlaAlaValSerLeuSerValSerThrIleSerGlyA 3443
 1130 ACACACTGCCCAAGATTAACCTGCAGCAGCTTCAGGAGGAGCTCCATC 1179
 3443 spAsnLeu.....IleAsnAlaAlaGluAlaGlySerAlaLeuThr 3456
 1180 TGCAGCGGGCTTCGATCGGCTTTACTGTGTATTAGGAAGAAATATCCAAC 1229
 3457 LeuSerGlyThrGlyThrAsnPheAlaThr..... 3466
 1230 TGATGGATCTGAATTTGTGTGCTGCGGATGGGAAGACAACTATAA 1279
 3467 ...GlyThrValThrValLeuLeuAsnGlyLys..... 3477
 1280 GTGGGTGCTTTAACGAGGTCAACAAAGTGTGCTCATCATCCACAGTC 1329
 3478 ..GlyTyThrAlaThrIleGlnSerAsnGlySerTrpSerValAsnVal 3493
 1330 GCTTTGGGGCCCTCTGAGCTCAAGAACTAGAGGAGCTGTCCAAATGAC 1379
 3494 ***AlaAlaAspValAlaAla.....LeuSe 3502
 1380 AGCAGGTTTACAGACATATGCTTCAGATCAAGTTTCAGAAACAATGCCCTCA 1429
 3502 rAspGlyThrSerTyThrValSerAlaSerAlaGlnAsp..... 3515
 1430 TTGATGCTTTTGGGCGCCCTTTCATCAGGAATGGAGCTGTCTCTCAGGCG 1479
 3516SerAlaGlyAsnSerAlaThrAlaSerArg 3525

541AsnAspIleAsnAsnIleIleSe 548
 1917 ATCAGTGAATGAAGAAACAGTACTTGGAACTACTGTAATGAGGAG 1966
 548 rGlyValAsnGlnAsnIleGly...AsnIleGlyThrAsnGlySerG 564
 1967 GTGCTGATGCTACTAAGATGACGGTCTCTACTCAAGTATTTCACA 2016
 564 lYglnIleSerLeuSerPheAspProIleThrGlyLysLeuThrThrAsp 580
 2017 TATGACACGAATGGTATACAGT...GTAAGAGTGGGGCTCTGGGAGG 2063
 581 PheAsnLeuAsnAlaSerIleSerGlyAlaThrPheArgSerLeuIleSe 597
 2064 AGTTAACGACGACGAGGAGATGATACCCAGCAGAGTGGGACACTGT 2113
 597 rThrThrSerArgArgSerThrPheIle..... 606
 2114 ACATACCTGGCTGGATTGAGATGATGAATACAAATGGAATCCACCAAGA 2163
 607AspAsnValMetGlyAsn..... 612
 2164 CCGTAAATTAATAGGATGATGTTCAACACAGCAAGTGTGTTTCAGCAG 2213
 613SerMetGlnSerPheAlaLeuAlaSe 621
 2214 ACATCTCGGAGGCTCATTTGTGGCTCTCATGTCCTCCAAT.....G 2257
 621 rSerSerLysSerGlnSerIleAlaMetSerGlnLysGlyAsnLeuTyra 638
 2258 CTCCATACCTGATCTCTCTCCACCTGGCCAAATCACCGACCTGAAGCG 2307
 638 laAspAlaSerAspTyr.....IleLysSer 646
 2308 GAAATTCACGGGGAGCTCTCATTAATCTGACTTGCAGACCTCTCGGGA 2357
 647 AspLeuAsnAsnGlySer..... 652
 2358 TGATTATGACCATGGAACAGCTCACAGTATATCATTCGAATAAGTACAA 2407
 653TyGlySerAsnLysGluHisSerLeuPheIleLeuProTyThrS 668
 2408 GTATCTCTGATCTCAGAGCAAGTTCATCAATCAATCTTCAATGATGAT 2457
 668 erSerGlnAsnValGluLeuSerLeuAsnGluGluSerLysGlyHisThr 684
 2458 ACTGCTCTCATC.....CCAAAGGAAGCCCACTCTCAGCA 2492
 685 LysGlyThrIleIleGlyTyrSerThrLeuLysAspSerGlyIleTyG1 701
 2493 AGTCTTT.....TTGTTTAAAC 2509
 701 yValTyAlaGlyTyrGluAspThrLysMetGlySerThrTyThrPheAsp1 718
 2510 CAGAAACATTAATCTTT..... 2526
 718 leAsnAsnArgThrTyTyAlaGlyLeuLysTyThrPheAsnThrLeuPhe 734
 2527GAAATGGCAGACATCTTTTCAATGCTATTTCAG.....GC 2561
 735 ThrThrGluLysGlyGlnGluValTyThrIleLysAlaGlnGlyLysAlaAl 751
 2562 TGTGATAAGTCAATCTGAATCAGAAATATCAAC 2598
 751 aLeuIleLysAsnAspLeuThrGluTyIleGlyAsn 763
 seq_name: pir2:B85547

seq_documentation_block:
 probable rtx family exoprotein [imported] - *Escherichia coli* (strain O157:H7)
 C:Species: *Escherichia coli*
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001


```

2305 GCGGAAATTCACGGGGCAGCTCTCATTAATCTGACTTGGACAGCTCTCTGG 2354
3770 LeuThrLeuAsnGlyLysThrTyrThrThrValGlnThrGlyGlySe 3786
2355 GGATGATTATGACCATTGACAGCAGCT.....C 2380
3786 rtrpSerTyrThrLeuGlySerAlaAspValThrAlaLeuAlaAspGlyA 3803
2381 ACAAGATATATCATTCGATATAGTACAGTATTCCTTGATCTCAGAGACAAG 2430
3803 snAlatyrValIleAsnAlaSerValSer.AsnAlaIleGlyAsnThrGI 3819
2431 TTCAATGAAT 2440
3819 ySerSerAsn 3822
_name: pir2:B71704
documentation_block:
1 surface antigen (sca3) RP451 - Rickettsia prowazekii
species: Rickettsia prowazekii
date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change
cession: B71704
ndersson, S.G.B.; Zomorodipour, A.; Andersson, J.O.; Sichert-
ure 396, 133-140, 1998
title: The genome sequence of Rickettsia prowazekii and the or
reference number: A71630; MUID:99039499
cession: B71704
tatus: preliminary; nucleic acid sequence not shown; translat
olecule type: DNA
esidues: 1-2340 <AND>
ross-references: GB:AJ235271; GB:AJ235269; NID:93868717; PIDN
xperimental source: strain Madrid E
enetics:
ene: sca3; RP451

```

```

alignment_scores:
  Quality: 143.50      Length: 809
  Ratio: 0.395         Gaps: 38
  Percent Similarity: 44.870  Percent Identity: 20.025

alignment_block:
  US-09-049-696-18 x B71704      ..

Align seg 1/1 to: B71704 from: 1 to: 2340

514 GGAGTATTTCGACGACTACAAATGATGAGAAATCTTACTATTATCCAAATGG 563
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
840 GlyAspIleAspPheAsnAsnIlySalaGlyLysPheIleLeuGlyAspGI 856
504 AAGAATA.....CAAGCAGTAGAAGTGTTCAGCAGGTATTACTGGTACAA 607
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
856 yAlaMetIleAspGlySerValLeuCysAsnGlyGlyValAlaGlyThr. 872
608 ATGTAGTAAAGAGTCTTCAGGAGGCGAGCTGTTACACCAAAAGATGCACA 657
    .. :.. :.. :.. :.. :.. :.. :.. :.. :.. :.. :..
872 ..... 872

653 TTCAATAAAGTAAACAGGACTCTATGAAAAGGATGTGAGTTTCTTCCCA 707
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
873 .....LeuAspPheIleGlyAs 878

708 ATCCGCCAGACGGAGAGGCTTCTATATGTTTGCACAACTGTTGATT 757
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
873 pGlyAsnValThrGlnAsnIleGlyAlaAspAsnAlaAsnSerIleSert 895

753 CTATAGTTGAATTCTGTACAGAACAAACCACAAAGAAAGCTCCAAAC 807
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
895 hrIle.....AsnIleGlnGlyAspAsn 902

808 AAGCAAAATCAAAAATGCAATCTCCGAAGCAGCATGGGAAGTGTACCGTGA 857
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```
903 ThrLys.....AsnValThrIleAlaAsnAspIlePheValAs 915
858 TTCTGAGGACTTTAAGAAACCACTCTCTATGACACACAGCCACCAATC 907
1145 ymetValPheLysHisGlnAspSerIleLeuGluLeu.....LysA 1159
1721 ACAGTCTGCAAGCAAGCTCACAAACCTTGACCTGACTGTACGTCCTCGT 1770
1159 snSerAsnThrAsnAspHisThrIleThrLeuThrSerAlaLeuAsp 1175
1771 GCGTCCAAATGCTACCTCCCTCCCAATACAGTCTGACTTCCAAAACGACAA 1820
1176 proGlyAsnAsnGlnPheGlyIleIleLysLeuIleThrAspThrAsnL 1192
1821 G.....GACACCAGCA 1831
1192 sLeuThrIleAspAsnAsnGlyAsnValAlaIleThrLeuGlyThrAla 1209
1832 AATTCCTCCAGCCCTCTGCTAGTTTATGCAAAATATTCGCCAAGGAGCCTCC 1881
1209 snHisMetLeuLysGlnLeuThrPheAlaSerIleAspAsnGlyAla... 1224
1882 CCAATTCACAGGCCAGTGTACAGCCCTGATTGATCATCATGATGGA 1931
1225 .....IleAlaLeuLysValGlyIleAsnValG 1234
1932 AACAGTTTACCTTGCAACTACTGGAT..... 1956
1234 uAsnValThrLeuAsnIleLysAspIleGluLeuAsnGluValAsnAla 1251
1957 .....AAT 1959
1251 snValLeuPheAsnLysAsnThrThrThrAlaThrGlyAsnIleAsn 1267
1960 GGA.....GCAGGCTCTGATCTACTAAGGATGA 1988
1268 GlyHisValAspPheGlnGlyAsnAlaGlyValIleAsnLeuAsnAsp 1284
1989 C...GGTCTCTACTCAAGGTATTTCAACAATATGACACAGAAATCGTAGAT 2035
1284 pIleGluIleAspGlySerValThrSerThrGlyAsnValAsnGly.... 1299
2036 ACAGTGTAAAAAGTCGGGCTCTGGGAGGAGTTAACGCGACGACGAGGA 2085
1300 ..ThrLeuAsnPheAsnGlySerGlyLysValThrGlyLeuIleAsnAsn 1315
2086 GTGATACCCACGACGAGTGAGCA.....CTGTACATACCTGG 2123
1316 IleValMetLeuGlnAlaGlyAlaGlyAspValSerLeuSerAlaSerG 1332
2124 CTGGATTGAGAAATGATGAATACAAATGGAATCCACCAAGA..... 2163
1332 yAsnTyrSerIleThrGluIleGlnGlyAsnGlyAsnAsnLeuThrP 1349
2164 .....CCTGAAATTAATAAGGATGATGTTTCAA 2190
1349 heAlaAlaAsnSerHisLeuThrThrAspIleAsnLysThrGlyGlyGln 1365
2191 CACAAGCAGTGTGTTTCAGCAGAACATCTCTGGGAGGCTCATTTGTGGC 2240
1365 AspLeuAsnLeuValPhe.....IleAsnGlyGlySerValSerG 1379
2241 TTCTGATCTCCCAATGTCCTCCATACCTGATCTCTTCCACCTGGCCAAA 2290
1379 ySerIleGlyAlaAsnAlaValGlyAspIleIleIleAsnAlaGlyS 1396
2291 TCACCGACCTGAAGCGGGAATTCAGGGGCGAGTCTCATTT..... 2331
1396 erValAsnPheSerAsnThrLeuLysSerGlyAsnIleValIleSerAsp 1412
2332 .....AATCTGACTTGGACA...GCTCCTGG 2354
1413 GlyAlaThrMetGlnValAsnAsnAsnValThrAlaThrAspIleSerG 1429
```

2355	GGATGATTATGACCATCGAACAGCTCACAGTATATATCATTCGCAATAAGTA	2404
	::: ::::	
1429	LysAsnAlaAsnGlyThrLeu.....	1437
2405	CYAAGTATTCTTGATCTCAGAGACAAGTTCAATGAATCTCTTCAAGTCAAT	2454
	: ::	
1438	..LysLeuAsnAsnHisThrProIleAsn	1446
2455	ACTACTGCCTCTATCCCAAAGGA.....GC	2480
	: :::	
1447	IleThrSerThrLeuGlyAsnAsnAlaIleGlyThrIleGluValAl	1463
2481	CAACTCTCAGGAAGCTCTTT.....TTGTTTAAACCAGAAAAAATCTACTT	2524
	: ::::	
1463	aAsnAsnAspValThrIleThrGlyThrLeuGlnAlaGlnAsnIleHisP	1480
2525	TTGAAAATGGCAGACAGACTTTTTCATTCTGCTATTTCAGTCTGTGTGATAAGTCT	2574
	: :: ::	
1480	heSerAsnAlaThrGln.....AlaAlaThrLeuThrLeu	1491
2575	GATCTGAAATCAGAATATCCAACATT	2601
	: ::::	
1492	GivAlaAlaSerGlnValThrAsnIle	1500

seq_name: pir2:D81411

seq_documentation_block:
probable lipoprotein C10629 [imported] - Campylobacter jejuni (strain NCTC 11168)

C; Species: Campylobacter jejuni
C; Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C; Accession: D81411
R; Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, T.; Whitehead, S.; Barrell, A.G.; et al.: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervirulence.
Nature 403, 665-668, 2000
A; Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervirulence.
A; Reference number: AB1250; MUID:20150912
A; Accession: D81411
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1-978 <PAR>
A; Cross-references: GB:AL111168; MID:g6967817; PIDN:CAB75265.1; PID:g696809
A; Experimental source: serotype O2, strain NCTC 11168
C; Genetics:
A; Gene: C10629

```
alignment_scores:      Length: 1010
                        Gaps: 46
                        Percent Identity: 19.010
                        Quality: 140.00
                        Ratio: 0.328
                        Percent Similarity: 42.277
```

alignment_block:

US-09-049-696-18 x D81411

Align seq 1/1 to: D81411 from: 1 to: 978

7 ACAGGGAGATGACAGCAATGGGGCCATTAAAGACTTCTGTGTTCAATCT 56
|||||:::|||||:::|||||:::|||||:::|||||:::
26 ThrGlySerIleTyPheGlyAsnAlaIlyIleSerPheValAs 42
|||||:::|||||:::|||||:::|||||:::|||||:::
57 G.....ATTCTTCACTTCTAGAGGGGGCCCTGAGTAAT..... 90
:::|||||:::|||||:::|||||:::|||||:::
42 nSerGlyThrIleLysSerLysGlnGlyAlaGlyValAsnIleSerGlnG 59
|||||:::|||||:::|||||:::|||||:::|||||:::
91TCACTCATTCACTGACCAACAATAGCTATGAGGCCATTCTCGTT 135
|||||:::|||||:::|||||:::|||||:::|||||:::
59 lyThrSerIleGluAsnPheAsnThrGly...ThrGlyIleIleGlu 74
|||||:::|||||:::|||||:::|||||:::|||||:::
136 GCATTCAGCCCATGTGCCAGAGATGAACACTCTATTCAACAATAAA 185
:::|||||:::|||||:::|||||:::|||||:::
75 GlyLysArgMetGlyIleValAsnValArgSerThrIleAsnThrPheValAs 91
|||||:::|||||:::|||||:::|||||:::|||||:::
186 GGACATGGTGACCCAGGCATCTCTGTATCTCTTTTCAAGCTCAGCAAGAC 235

[illegible]

304 uGlyGlyArgGlyValGlyIleLeuAsnArgSerGlyLysIleGluGlys 321
 1028 TCAGACAGATTGAGTGGGTCTGGGT 1056
 321 erileysValGluAspGlyAlaThrValThrAlaThrSerAsnArgAla 337
 1057GGGATGGTGACA 1068
 338 IleAlaAsnSerGlySerGlySerIleThrGlyIleThrValSerGI 354
 1068 1068
 354 yLysAsnThrLysLeuGluGlyAsnIleLeuAsnThrGlyAsnAlaSerI 371
 1069TTTGACAGTGTGCCCATGTACAAAGTGAATC 1101
 371 leGlySerAspIleLysIleGluGlyGlyAlaLysValGluGlyLeu 387
 1102ATACAGATAACAGTGG 1118
 388 ValAsnGlnGlyAsnGlySerIleSerGlySerValGlnValSerGlyI 404
 1119 CAGTGACGGACACACTGCCAAAGATTACTCGACGAGCTTCAGGAG 1168
 404 ySerSerIleAspSerIleThrAsnGluGlyAsnGlyAlaIleSerGlys 421
 1169 GGAGTCCATCTGCAGCGGCTTCGATCGGCATTACTGTGATT 1212
 421 erIleThrValTyLysAspSerLysLeuAspSerIleThrAsnThrSer 437
 1213 AGAAGAAATATCCAACTGATGATCGAAATGTCTGCTGACGGATGG 1262
 438 ThrSerSerThrGlyIleSerGlySerIleThrAsnAs 450
 1263 GGAGACACACTATAAGTGGTGTTTACAGAGGTCAACAAAGTGGT 1312
 450 nSerAspAsnLysLeuGluIleSerAsnSerGlyA 462
 1313 CCATCATCCACACAGTGGCTTTGGGGCCCTCTGCAGCTCAGAACTAGAG 1362
 462 snlle 463
 1363 GAGCTGTCCAAATGACAGCAGGTTTACAGACATATGCTTCAGA 1406
 464GlyGlyLysIleGluSerThrGlySerAlaAs 474
 1407 TCAAGTTCAGAACAAATGGCTCATTTGATGCTTTTGGGCC 1447
 474 pMetValIleSerAsnSerAsnGlyGlyThrIleSerGly...GlyIle 490
 1448 TTTTCATCAGNAATGGAGTGTCTCTCAGCGCTCCATCCAGCTTGAGAGT 1497
 490 erSerSerGlySerGlySerThrIleSerAsnSer 502
 1498 AAGGATTAAACCTCCAGACAGCAGCTGGATGAATGGCACAGTGATCGT 1547
 503 GlnGlySerThrIleAsnAsnGlyIleThrValSerGlySerAlaGlnVa 519
 1548 GGACAGCACCGTGGAAAGGACACTTTGTTCTTATCA 1585
 519 lGluIleSerAsnGlnGlySerValGlyLysAsp 530
 1586 CTGGACAAACGACGCTCCCAATCTCTCTGGGATCCAGTGGACAG 1635
 530 530
 1636 AAGCAAGTGGCTTTGTAGTGGCAAAACACCAAAATGGCTTACCTCCA 1685
 531GluAsnGlyAsnThrValThrAsnAsn 539
 1686 AATCCAGGATTCGTAGGTGGCACTTGGAAATACAGTCTGC 1729
 540GlySerGlySerValGlyIleLysAspTrpLeuValSerThrA 554

1730 AACCAAGCTCACAAACCTTGACCTGACTGTACGTCCCGTGGG 1773
 554 spLysAsnThrGlyLysLeuAsnThrValValIleGlySerArgAla 570
 1774 TCCAACTGCTACCTCCATTCATACAGTCTCCAAACGAAACAAAGGA 1823
 571 PheAsnValLysValGluAsnIleThrVal...AspGlnSerAsnValAs 586
 1824 CACCAAGCAAAATTCGCCAGCCCTCTGTAGTTTATGCAAAATATTGCGCAAG 1873
 586 pLeuGluGluLeu 590
 1874 GAGCCTCCCAANTCTCAGGCCAGTCTCACGCCCTGATTGAATCAGTG 1923
 591AsnAspIleAsnAsnIleIleSerGlyVal 600
 1924 AATGGAAAAACAGTTACTTTGGAACTACTGGATAAATGGAGCAGGTGCTGA 1973
 601 AsnGlnAsnAsnIleGly...AsnIleGlyThrAsnGlySerGlyGluI 616
 1974 TGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCACAACATTATGACA 2023
 616 eSerLeuSerPheAspProIleThrGlyLysLeuThrThrAspPheAsnL 633
 2074 CGAATGCTAGATACAGT...GTAAAAGTGGCGCTCTGGGAGGAGTTAAC 2070
 633 euAsnAlaSerIleSerGlyAlaThrPheArgSerLeuIleSerThrThr 649
 2071 GCAGCCAGCAGGAGTGTATACCCACAGAGTGAGCAGCTGTATACATACC 2120
 650 SerArgArgSerThrPheIle 656
 2121 TGGCTGGATTGAGAATGATGAATACATGGAATCCACCAGACCTGAAA 2170
 657AspAsnValMetGlyAsn 662
 2171 TTAATAAGGATGATGTTCAACACAGCAAGTGTGTTTCAGCAGAACATCC 2220
 663SerMetGlnSerPheAlaLeuAlaSerSerSer 673
 2221 TCGGGAGGCTCATTTGTGGCTTCGTGATGTCCAAATGCTCCCAT 2264
 674 LysSerGlnSerIleAlaMetSerGluLysGlyAsnLeuTyAlaAspAl 690
 2265 ACCTGATCTCTCCACCTGGCCAAATCACCAGCTGAAGGGGAAATTC 2314
 690 aSerAspTyIleLysSerAspLeuA 699
 2315 ACGGGGGCAGTCTCATTAATCTGACTTGGACAGCTCTGGGGATGATTAT 2364
 699 snAsnGlySerTyr 703
 2365 GACCATGGNACAGCTCACAACTATATCATTCGAATAAGTACAAGTATTCT 2414
 704 GlySerAsnLysGluHisSerLeuPheIleLeuProTyThrSerSerGI 720
 2415 TGATCTCAGACACAAGTTCATGAATCTCTTCAAGTGAATACTACTGCTC 2464
 720 nAsnValGluLeuSerLeuAsnGluLysSerGlyHisThrLysGlyT 737
 2465 TCATCCCAAAGGAAGCAACTCTGAGGAAGTCTTT 2499
 737 hrIleIleGlyTySerThrLeuLysAspSerGlyIleTyGlyValTy 753
 2500TTCCTTTAAACACAGAAA 2516
 754 AlaGlyTy-GluAspThrLysMetGlySerThrTyThrPheAspIleAsnAs 770
 2517 CATTACTTTTG 2527
 770 nArgThrTyTyAlaGlyLeuLysTyPheAsnThrLeuPheThrThrG 787

2528 A A A T G C C A C A G A C T C T T T C A T T G C T A T T C A G G C T G T T G A T 2568
787 I u L y S G l n G l n u V a l T r y I l e L y s A l a G l n L y s A l a A l a L e u l l e 803
2569 A A G T G C G A T C T C A A A T C A G A A A T A T C C A A C 2598
804 L y s A s n A s p L e u T h r G l u L y s I l e G l y A s n 813

seq_name: pir2:A43855

seq_documentation_block:
high-molecular-weight surface-exposed protein HMW1 - Haemophilus influenzae
C:Species: Haemophilus influenzae
C:Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999
C:Accession: A43855
R:Barenkamp, S.J.; Leininger, E.
Infect. Immun. 60, 1302-1313, 1992
A:Title: Cloning, expression, and DNA sequence analysis of genes encoding nontypeable Haemophilus influenzae type A pertussis toxin.
A:Reference number: A43855; MUID:92192797
A:Accession: A43855
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 11536 <BAR>
A:Cross-references: GB:U008976; GB:M84616; NID:9475770; PIDN:AA20527.1; PID:g475771
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:89235, NCBIP:89239)

alignment_scores:		
Quality:	138.50	Length: 936
Ratio:	0.318	Gaps: 49
Percent Similarity:	46.581	Percent Identity: 20.406

alignment block:

US-09-049-696-18 x A43855

Align seq 1/1 to: A43855 from: 1 to: 1536

217	TTTGAAGCTACA.....CGAAAGCGATTTTTATTTCAAAAATGCT	254
	:	:
631	PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnIle	643
255	TGCCATTGTTGATTCTCGAAACATGAAGACAAAGCGTGAATGATGAGAC	304
	: :	: :
643	eserMetValLeuPro.....LysAsnGluSerGlyTyrAspLysP	657
305	CAAAATCTTGAGACCTACAAAAATGCTGATGTTCTG...GTTCTGAGCTG	351
	: :	: :
657	heLysGlyArgThrTyrTrpAsnLeuThrSerLeuAsnValSerGluSer	673
352ACTCTCCAGGTAAATGAT.....	369
674	GlyGluPheAsnLeuThrIleAspSerArgGlySerAspSerAlaGlyTh	690
	: :	: :
370GAACCCCTACACT.....	388
	: : :	: : :
690	rIeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThr	707
389	TGGGCAACTGTGGAGAGA.....AGGCTGAAAGGATCCACTCACTCCT	432
	: :	: :
707	hcPheAsn.ValGluArgAsnAlaArgValAsnPheAspIleLysAlaPr	723
433	GATTTCATTGCAGGAAAAAAGTTACTGAAATATGACACCAAGGTAGGGC	482
	: : :	: : :
723	oileGlyIleAsnLysTyrSerSerLeuAsnTyrAlaSerPheAsnGlyA	740
483	ATTTCCTCATGATGGGCTCATCTACGATGGGCGATATTGTGACGAGTACA	532
	: : :	: : :
740	snIleSerValSerGly.....GlyGlySerValAspPheThr	752
533	ATAATGATGAGAAATTTACTTTATCCAATGGAAAGATACA.....AGCA	576
	: : :	: : :

LeuLeuAlaSerSerSerAsn.ValGlnThrProGlyValValIleAsnS 769
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487

1237	TCGTAAATGTCCTGCTGACGGATGGGGAACACACACTATAAGTGGG...	1284
1059		1072
1285		1089
1072		1089
1313	CGATCATCCACACAGTCGCTTTGGGCGCCTCTGCAGCTCAAGAACTAGAG	1362
1089	laAspGlyHisLysValThrLeuHis	1097
1363	GAGCTCTCAAAATG...ACAGGAGTTTACAGACATATGCTTCAGATCA	1409
1098		1112
1410	AGTTCAGAAACATGGGCTCATTGATGCTTTTGGGCGCCTTTTCATCAGGAA	1459
1112	rSerAspAsnAsnAla	1117
1460	ATGGAGCTGTCTCTACGCGCTCCATCCAGCTTGAGAGTAAGGATTACCC	1509
1118		1127
1510	CTCCAGAACAGCCAGTGATGATGACAGTGCCTGGACAGCACCGCT	1559
1128	ValAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe	1144
1560	GGGAAAGGACACTTTGTTCTTATCATCCCTGGACAGCACGCGCTCCCAAA	1609
1144	rGlyGlu	1151
1610	TCCTTCTCTGGATCCAGTGGACAGCAAGGTGGCTTCTAGTGGAC	1659
1152		1156
1660	AAAAACACCAAAATGGCTTACCTCCAAATCCAGGCAATTGCTAAGTTGG	1709
1157	AlaThrThrGly	1169
1710	CACHTGGAATACAGTGTGCAACAGCTCACAACTTGACCCCTCAGTCG	1759
1169	ySerIleLeuGlyGlyIleGluSerSerGlySerValThrLeuThra	1186
1760	TCAGTCCCGTGGC...TCAATGCTACCTGCGCTCCCAATTACA	1800
1186	laThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr	1202
1801	GTGACTTCCAAACGACACAGGACACGAC	1830
1203	ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLy	1219
1830		1830
1219	sGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyClyt	1236
1831		1854
1236	hrIleSerGlyThrValGluValLysAlaThrGluSerLeuThrThr	1252
1855	TATGCAATATATCGC...CAAGGAGCCCTCCCAATCTCCAG	1892
1253	GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe	1269
1893	GGCCAGTCTCACAGCCCTGATGATGATGATGATGATGATGATGATGATGAT	1942
1269	rAlaThrGlyThr...IleGlyGlyThrIleSerGlyAsnThrValAsnV	1285
1943	TGGAA...CTACTGGATATGAGGAGGCTGCTGAT	1974
1285	alThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn	1301

1975	GCTACTAAG.....GATGACGGTGTCTACTCAAG	2003
1302	AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh	1318
2004	GTATTTCAACACTTATGACACGAATGGTAGATACAGTGTAAAGTCGGG	2053
1318	rGluAlaSerSerHisIleThrSerAlaLysGlyGlnValAsnLeuSerA	1335
2054	CT.....CTGGGAGGAGTTAAACGACGACGACGAGAGGAGTG	2088
1335	laGlnAspGlySerValAlaGlySerIleAsnAlaAla.....AsnVal	1349
2089	ATACCCAGCAGAGTGGAGCAGCTACATACCTGGCTGGTGGATTCAGATGA	2138
1350	ThrLeuAsnThrThrGlyThrLeuThr.....ValLysGlySe	1363
2139	TGAATACATGGAATCCACCAAGACCTGAAATTAATAGGATGATGTTTC	2188
1363	rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG	1380
2189	AA.....CACAAAGCAAGTGTGTTTTCAGCAGA	2214
1380	luLeuAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn	1396
2215	ACATCTCGGGAGGCTCATTTCTGGCTTCTGATCTCCCAATGCTCCCAT	2264
1397	AlaAsnGlySerGlySerValIleAlaThrThrSerSerArgValAsnIl	1413
2265	ACCT...GATCTCTCCACCTGGCCAAATC.....ACCG	2296
1413	ethrGlyAspLeuIleThrIleAsnGlyLeuAsnIleIleSerLysAsnG	1430
2297	ACCTGAAGCGGAAATTCACGGGGCAGCTCTCATTAATCTGACTTGGACA	2346
1430	lyIleAsnThrValLeuLeuLysGlyValLysIleAspValLysIle	1446
2347	GCTCCTGGG...GATGATTATGACCATGGAACAGCTCACAGTATATCAT	2393
1447	GlnProGlyIleAlaSerValAspGluValIleGluAlaLys.....	1460
2394	TCGAATAAGTACAAGTATTTGATCTCAGACAGCAAGTTCAATGATCTC	2443
1461	ArgIleLeuGluLysValLysAspLeuSerAspGluArgGluAlaL	1477
2444	TT 2445	
1477	eu 1477	

seq_name: pir2:T21810

seq_documentation_block:
hypothetical protein F35G12.4b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21810
R:Chui, C.
submitted to the EMBL Data Library, October 1994
A:Reference number: Z19473
A:Accession: T21810
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-683 <WIL>
A:Cross-references: EMBL:Z46242; PIDN:CAA86338.1; GSPDB:GN00021; CESP:F35G12.4b
A:Experimental source: clone F35G12
C:Genetics:
A:Gene: CESP:F35G12.4b
A:Map position: 3
A:Introns: 62/3; 105/1; 135/3; 200/3; 269/3; 332/2; 390/1; 425/3; 458/3; 490/1; 542/2

alignment_scores:
Quality: 138.00 Length: 602
Ratio: 0.531 Gaps: 33

Percent Similarity: 43.189 Percent Identity: 20.598

alignment_block:

US-09-049-696-18 x T21810 ..

Align seg 1/1 to: T21810 from: 1 to: 683

```
1370 AACACTAAGTGGTCTTTACGAGGTCAACAAAGTGGTGCCATCAT 1319
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
172 AsnAsnLeuIleGlyCys.....LysAspSer.....I 181
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1320 CCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAAGCTA..... 1359
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
181 eTySerLeuAlaThrThrProAsnLeuSerLeuValLeuGlyAlaGlyT 198
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1360 .....GAGGAGCTGTCC 1371
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
198 hrGluLysCysIleArgLeuPheAspProArgThrAsnGluLysIleMet 214
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1372 AAATGACAGAGGTGTACAGACATATGCTTCAGATCAAGTTCAAGACAA 1421
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
215 LysLeuArgGlyHisThrAspAsnValArgAlaLeuValValAsnAspAs 231
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1422 TGGCCTCATTTGCTTTTGGGGCCCTTTCATCAGGAAATGGAGCTCTC. 1470
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
231 pGlyThr.....ArgAlaLeuSerAlaGlySerAspAlaThrI 244
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1471 .....TCTCAGGCTCCATCCAGCTTGAGAGTAAG 1500
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
244 leArgLeuTTPAspIleGlyGlnArgCysIle.....AlaThr 257
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1501 GGATTAACTCCAGACACAGCTGATGAATGGACAGTATCGTGA 1550
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
258 CysIleAlaHisGluGluGlyValTrp.....ThrLeuGlnValAs 271
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1551 CAGC.....ACCGTGGAAAGGACACTTTGTTCTTA 1582
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
271 pSerSerPheThrThrValTrpSerAlaGlyLysAspLysMetValVal 288
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1583 TCACC.....TGGCAACGACGCTCCCAAAATCTCTCTGGGATCCC 1626
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
288 ySThrProLeuTyPAspPheThrLysSerGlnLeuLeuPheLysGluGlu 304
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1627 AGTGCAGACAGCAAGGTGGCTTTGTAGTGCACAAAACACCAAAATGCC 1676
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
305 AlaProValLysLys.....LeuLeuLeuSe 313
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1677 CTACCTCCAAATCCAGGCATTTGCTAAGTTGGCACTTGGAAA..... 1719
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
313 rGluLysAspAsnProValSerLeuTTPValGlyThrTrpLysSerAspI 330
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1720 .....TACAGTCTCAAGCAAGCTCAACAACTTGACCCGTGACGTGC 1761
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
330 leLysArgTrpSerIleArgProSerAlaGln.....LeuSerIle 343
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1762 AGTCCCGCTGGTCCCAATGTACCTGCCCTCCCAATTACAGTACTCCAA 1811
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
344 GlyGlyAspGluAspGlyProSer.....ThrSerAs 354
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1812 AACGAACAAGACACACAGCAAAATTCGCCAGCCCTCTGGTAGTTATGCAA 1861
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
354 nAlaAsnHisSerValSerAlaSerSerProProValThrPheLysT 371
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1862 ATATTCGCCAAGGAGCCTCCCAATTTCTCAGGGCCAGTGCACAGCCCTG 1911
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
371 yTrIleArg..... 373
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1912 ATTGAATCAGTGAATGAAGAAACAGTACCTTGGTGAATATGG 1961
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
374 ValLysAspGlnLysGlyGlnGlnSerThrProGluLeuValIle..... 388
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1962 AGCAGGTGCTGCTACTACTAAGGATGACGGTGTCTACTCA.....AGGT 2005
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

```
389 .ProGlyAlaProAlaIleLysLysHisAlaMetLeuSerAspLysArgH 405
2006 ATTTCAACACTTATGTACACGAATGGTAGATACAGTGTAAAGATCGGGCT 2055
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
405 isValLeuThrArgAspSerAspGlyAsnVal.....Ala 416
2056 CTGGGAGGAGTTAAAGCAGCAGCAGCG..... 2082
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
417 LeuTyPAspValLeuAlaAlaArgLysIleLysAspTyPArgLysArgI 433
2083 .....AGATGATATACCCAGCAGAGCTGGAGCAGCTGTACATACCTGGCT 2125
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
433 ePheGluGluValValAspGluAsnSerArgGlnValTyPTrIleProSerT 450
2126 GG.....ATTGAGAAATGATCAA 2142
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
450 rpPheValValAspSerLysSerGlyMetLeuGlnIleThrLeuAspGlu 466
2143 ATA.....CAATGGAATCCACCAAGACCTGAA...ATTAAATA 2177
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
467 LeuAspAlaLeuSerSerTrpLeuSerSerLysAspAlaGlyPheAspAs 483
2178 GGATGATGTTCAACACAAAGCAAGTGTGTTTCAGCAGAACATCCCTCGGAG 2227
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
483 pAsnAspArgGluThrLysGlnIleIle.....ValAsnTyPArgGly 497
2228 GCTCATTTGTGGCTTCTGATGTCCCAATGCTCCCATCTACCTGATCTCTTC 2277
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
497 LyMetMetLeuArgSer.....LeuPheGluArgTrp 507
2278 CCACCTGGCCAAATCACCAGCCTGCAAGCGGAAATTCACGGGGCAGTCT 2327
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
508 ProProCysLysMetThrAsnValAspAla..... 517
2328 CATTAATCTGACTGGACAGCTCTCGGGATGATTATGACCATGGAACAG 2377
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
518 .....AlaAspAlaAspAspValGlnLysAlaThr. 527
2378 CTCACAAGTATATCATTCGAATAAGTACAAGTATCTTGATCTCAGAGAC 2427
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
528 .....LeuAsnPheIleSerLeuProGlu 535
2428 AAGTTCAATGAATCTCTCAAGTCAATACTACTCTCATCCCAAGGA 2477
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
536 His.....ThrProLeuIleIleCysG 543
2478 AGCCAACCTCTGAGGAAGTCTTT.....TTGTTTAAACCCAGAAAACATTA 2521
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
543 uGlyAsnGlyArgProLeuTyPArgLeuLeuValGlyAspAlaGlyLysG 560
2523 CTTTTGAA...AATGGCACAGATCTTTTCATGTTGCTATTCAGGCTGTGAT 2568
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
560 luPheGluAlaAsnGluLeuAlaGlnIleAlaProMetTrpValIleAsp 576
2569 AAGTCTCATCTGAATCAGAAATCCAAACATTCACAGCAGTATCTTTGTT 2618
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
577 AlaIleGlu...ArgAsnGlnLeuProLysPheAsnLysMetProPheTy 592
2619 TATT.....CCTCCACAGACTCCGCCAGACAGACACCTAGTCTCGATGAAA 2662
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
592 rLeuLeuProHisProSerThrAsnProLysGlnProLysLysAspArgL 609
2663 CGTCTGCT..... 2670
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
609 euSerAlaThrGluMetLeuGlnValLysLysValMetGluHisValTyP 625
2671 .....CCTTCTCTCTAATAT 2684
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
626 GluLysIleLeuSerThrAsnAspValGlySerIleProLeuAsnGlnI 642
2685 TCATATCAACAGCACCATT.....CCTGGCA 2710
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
642 eHisThrLysMetGluMetTyPAsnAspGlnArgLeuGluProAspM 659
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

1583 TCACC.....TGGACAAGCAGCCTCCCAATCTCTCTGGGATCCC 1626
1584 ysthrProLeuTyrAspPheThrLysSerGlnLeuLeuPheTyrGlu 304
1585 AGTGGCAGAGCAAGTGGCTTTGTAGTGGACAAACACCAAAATGCG 1676
1586 AlaProValLysLys.....LeuLeuLeuSe 313
1587 CTACCTCCCAATCCAGGCATTGCTAAGTTGGCACTTGGAAA..... 1719
1588 rGluLysAspAsnProValSerLeuTrpValGlyThrTrpLysSerAspI 330
1589TACAGTCTGCAAGCAAGCTCAACAACTTACCCCTGACTGTC 1761
1590 leLysArgTTrpSerIleArgProSerAlaGln.....LeuSerIle 343
1591 AGTCCCGTGGTCCCAATGTACCTGCCCAATTACAGTACTGACTCCAA 1811
1592 GlyGlyAspGluAspGlyProSer.....ThrSerAs 354
1593 AACGAACAAGGACACCAAGCAATTCGCCAGCCTCTGTTAGTTATGCAA 1861
1594 nAlaAsnHisSerValSerAlaSerSerProProValThrPheLysT 371
1595 ATATTCCCAAGGAGCCTCCCAATTCCTCAGGCCAGTGTACAGCCCTG 1911
1596 yrIleArg..... 373
1597 ATTAATCAGTGAATGAAACAAAGTTACCTTGGAACTACTGTGATAATGG 1961
1598 ValLysAspGlnLysGlyGlnSerThrProGluLeuValIle..... 388
1599 AGCAGGTGCTGATGCTACTAGGATGAGGTGCTACTCA.....AGGT 2005
1600 ProGlyAlaProAlaIleLysLysHisAlaMetLeuSerAspLysArgH 405
1601 ATTTCAACAATTATCACACGAATGTAGTACAGTGTAAAAGTCGGGCT 2055
1602 isValLeuThrArgAspSerAspGlyAsnVal.....Ala 416
1603 CTGGAGGAGTTAAGCAGCAGCAGCG..... 2082
1604 LeuTyrAspValLeuAlaAlaArgLysIleLysAspTyrGlyLysArgII 433
1605 AGAGTGATACCCAGCAGAGTGAGGAGCTGTACATAGCTGGCT 2125
1606 ePheGluGluValValAspGluAsnSerArgGlnValTyrIleProSerT 450
1607 GG.....ATTGACAATGATGAA 2142
1608 rPheValValAspSerLysSerGlyMetLeuGlnIleThrLeuAspGlu 466
1609 CAATGGNATCCCAAGACCTGAA...ATTAATAA 2177
1610 LeuAspAlaLeuSerSerTrpLeuSerSerLysAspAlaGlyPheAspAs 483
1611 GGATGATGTTCACACAAAGTGTGTTCACAGACAACATCCTCGGAG 2227
1612 pAsnAspArgGluThrLysGlnIleIle.....ValAsnTyrGlyG 497
1613 GTCATTTGTGGCTTCTGTATGTGCCAAATGCTCCCATACCTGATCTCTC 2277
1614 LeuMetLeuArgSer.....LeuPheGluArgTrp 507
1615 CCATCGGCGCAATCACCGACCTGAAGCGGAAATTCACGGGGCAGTCT 2327
1616 ProProCysLysMethThrAsnValAspAla..... 517
1617 CATTAACTGACTTGGACAGCTCCTCGGGGATGATTATGACCATGGAACAG 2377
1618 AlaAspAlaAspValGlnLysAlaThr. 527
1619 CACCAAGTATATCATTCGAATAAGTACAAGTATTCTTCTGATCTCAGAGAC 2427

2711 TTCAC.....ATTTTAAATATATGTGAAGTGCATAGGAGAACTG 2751
2712 etAspLeuArgThrValLysHisLeuTyrTrpLysGlnSerGlyGluLeu 675
2752 CAGCTG 2757
676 LeuLeu 677
seq_name: pir2:T21808
seq_documentation_block:
hypothenical protein F35G12.4a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21808
R:Chui, C.
submitted to the EMBL Data Library, October 1994
A:Reference number: 219473
A:Accession: T21808
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-686 <WIL>
A:Cross-references: EMBL:Z46242; PIDN:CAA86335.1; GSPDB:GN00021; CESP:F35G12.4a
A:Experimental source: clone F35G12
C:Genetics:
A:Gene: CESP:F35G12.4a
A:Map position: 3
A:Introns: 62/3; 105/1; 135/3; 200/3; 269/3; 332/2; 390/1; 425/3; 458/3; 490/1; 542/2; 686/3
alignment_scores:
Quality: 136.50 Length: 536
Ratio: 0.559 Gaps: 30
Percent Similarity: 45.522 Percent Identity: 21.269
alignment_block:
US-09-049-696-18 x T21808 ..
Align seg 1/1 to: T21808 from: 1 to: 686
1270 AACATATAAGTGGTCTTACAGAGTCAACAAAGTGGTCCCATCAT 1319
1271 AsnAsnLeuIleGlyCys.....LysAspSer.....11 181
1320 CCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAACTA..... 1359
1321 eTyrSerLeuAlaThrThrProAsnLeuSerLeuValLeuGlyAlaGlyT 198
1360GAGGAGCTGTCC 1371
1372 hrGluLysCysIleArgLeuPheAspProArgThrAsnGluLysIleMet 214
1373 AAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGAACAA 1421
1374 LysLeuArgGlyHisThrAspAsnValArgAlaLeuValValAsnAspAs 231
1422 TGGCCTCATGATGCTTTGGGGCCCTTTTCATCAGGAATGGAGCTGTC. 1470
1423 pGlyThr.....ArgAlaLeuSerAlaGlySerAspAlaThrI 244
1471TCTCAGCGCTCCATCCAGCTTGAGAGTAAG 1500
1472 leArgLeuTrpAspIleGlyGlnArgCysIle.....AlaThr 257
1501 GGATTAACCTCCAGAACAGCAGTGGATGAATGGCAGCTGATCGTGA 1550
1502 CysIleAlaHisGluGluGlyValTrp.....ThrLeuGlnValAs 271
1551 CAGC.....ACCGTGGGAAGGACACTTTGTTCTCTTA 1582
1552 pSerSerPheThrThrValTyrSerAlaGlyLysAspLysMetValVal 288


```

528 .....LeuAsnPhelSerLeuProGlu 535
2428 AAGTTCATGAATCTCTCAAGTGAATCTCTCATCCCAAGGA 2477
536 His.....ThrProLeuIleIleCysG1 543
2478 AGCCAACTCTGAGAACTCTT.....TTGTTTAACCCAGAAAACATTA 2521
543 uGlyAsnGlyArgProLeuTyrArgLeuLeuValGlyAspAlaGlyLysG 560
2522 CTTTGTAA...AATGGCACAGATCTTTTCATTGCTATTACAGGCTGTGAT 2568
560 lUpHeGluAlaAsnGluLeuAlaGlnIleAlaProMetTTrpValIleAsp 576
2569 AAGTGCATCTGAATCAGAAATATCCAACTTCACGAGTATCTTTGTT 2618
577 AlaIleGlu...ArgAsnGlnLeuProLysPheAsnLysMetProPheTy 592
2619 TATT.....CCTCCACAGACTCCGCCAGACACACCTAGTCTCGATGAAA 2662
592 rLeuLeuProHisProSerThrAsnProLysGlnProLysLysAspArgL 609
2663 CGTCTGCT 2670
609 euSerAla 611
seq_name: pir2.A85547
seq_documentation_block:
hypothetical protein 20609 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: A85547
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85547
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1461 <STO>
A:Cross-references: GB:AE005174; NID:g12513364; PIDN:AAG54837.1; GSPDB:GN00145; UWGP:206
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 20609

```

```

alignment_scores:
  Quality: 136.50      Length: 758
  Ratio: 0.392        Gaps: 40
  Percent Similarity: 45.910  Percent Identity: 21.108

```

```
alignment_block:
```

```
US-09-049-696-18 x A85547 ..
```

```
Align seg 1/1 to: A85547 from: 1 to: 1461
```

```

892 ACACAGCCACCAATCCACCTCTCA.....TT 920
197 ThrAspSerThrThrProThrPheSerGlySerGlyGlnAlaGlyAlaThr 213
921 GCTGCAGATT.....GGACAAAGAAATGTGTGTTAGTCCCTTG 958
213 rIleGlnIleLysAspSerAsnGlyAsnThrIleAlaSerThrGlnVala 230
959 ACAAACTTGA.....AGCATGGCGACTGTGTACCGCCTCAAT 996
230 spAsnAsnGlyHisTrpSerValSerLeuProThr..... 241
997 CGACTGAATCAAGCAGCCAGCTTTTCCCTGCTGCAGACAGTTGAGCTGGG 1046

```

```

242 .....GlnSerAlaGlyGluHisThrTrpSerValValGlnIleValG1 256
1647 GTCTCGGTT.....GGGATGGTGACATT.....GACAGTG 1078
255 ySerThrIleThrAspAlaGlySerIleThrLeuThrIleAspAsnSerG 273
1079 CTGCCCATGTACAAAGTGAACATACAGATAACACAGTGCAGTGACAGG 1128
273 lNAlaSerValGln.....ValAlaThrThrAlaGlyAspAsnIle 286
1129 GACACACTCGCCAAAGATTACCTGACAGCAGCTTCAGGAGGAGCTCC... 1176
287 lLeAsnAlaSerGluGlnAlaAlaGlyPheThrLeuSerGlyThrSerSe 303
1177 ....ATCTGCAGCGGCTTCGATCGCGATTCTACTGCTAGTAGGAAGAAT 1222
303 rHisLeuAlaGlnGlyThrGluLeuThrValThrLeuAsnGlyLysThr 320
1223 ATCCAAC..... 1230
320 yThrThrSerValGlyAlaAsnGlyAlaThrPpSerValGlnValProThr 336
1231 .....GATGGATCTCAAAATTGTGCTGCTGACG... 1257
337 AlaAspAlaGlnAlaLeuGlyGluGlyAsnGlnAlaValLeuValSerG1 353
1258 ....GATGGGAAGACAACTATAAGTGG..... 1284
353 yLysAspAlaThrGlyAsnThrValThrGlyAlaGlnLeuLeuThrVala 370
1285 .....TGCTTTAACGAGGTCAAAACAAAGTGGTCC 1314
370 spThrGlnProThrLeuAlaIleAsnThrIleAlaGlnAspAsnIle 386
1315 ATCATC.....CACACAGTCGCTTG.....GGCCCTCTGTC 1346
387 lLeSerAlaAlaGluHisAsnValAlaLeuValLeuSerGlyThrSerAs 403
1347 AGCTCAAGAAGTACAGAGCTGTCCAAATACAGACAGAGGT.....T 1387
403 nAlaGluAlaGlyGlnThrValThrLeuThrValAsnGlyLysSerHis 420
1388 TACAGACATATGCTTCAGAT..... 1407
420 hrAlaThrValGlySerAspGlyThrTrpGlnValThrLeuProAlaThr 436
1408 CAAGTTCAGAACAAATGCCTCATTCATGCTTTGGGCCCTTTCATCAGG 1457
437 GluValGln.....AlaLeuAlaGluG1 444
1455 AATGGAGCTGCTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAA 1507
444 yAsnTyrAlaValAsn...AlaSerValSerAspArgAlaGlyAsnThr 460
1508 CCCTCCAGAACAGCCAGTGGATGAATGGCACA.....GTGATC 1545
463 hrSerHisSerAlaAsnPheThrValAspThrSerAlaProValValSer 476
1546 GTGCACAGCACCGTGGGAAAGGACACTTTGTTTCTTATCACCTGGACAAC 1595
477 ValAsnThrValAlaGlyAspAspIleLeu.....As 487
1596 GCAGCTCCCAAAATCCTTCTCGGATCCAGTGGACAGAACGAAGGTG 1645
487 nAsnAlaGluGlnAlaValAlaGlnIleIleSerGlyGlnValSerGlyA 504
1646 GCITTTGAGTGGACAAAACACC...AAATGGCTACCTCCAATCCCA 1692
504 lAserProGlyAspThrValThrValLysLeuGlyThrHisValLeuThr 520
1693 GGATTCCTAAGTT...GGCAGCTTGGAAATACAGTCTGCAAGCAAGC... 1737
521 GlyIleValLeuAlaAspGlySerTrpAsnValAlaLeuAspProAlaVa 537

```

```
1738 .....TCACAACCTTGACCTGACTGTCAGT 1765
537 lThrArgThrLeuAspArgGlyAlaAsnThrIlePheValThrValThrA 554
1766 CCGTGCCTCAATGCT.....ACCGTG..... 1788
554 sPalaAlaGlyAsnThrGlyAlaAlaSerArgAlaIleThrLeuValGly 570
1789 .....CCTCAATTACAGTGTTCCTCAAAACGAACAAAGGACACCAAGAA 1832
571 ValSerProLeuIleThrIleAsnThrValSerGlyAspAsp..... 584
1833 ATCCCCACCCCTCTGTTAGTTATGCAATATTCGCCAAGGAGGCTCC 1882
585 .....IleIleSerGlyAlaGluLysGlyAlaProL 595
1883 CAATTCACAGGCCAGTGTACAGCCCTGATTGAATCAGTGAATGAA 1932
595 euThrLeuThrGlySerThrGln.....Gln**GluThrGlyGln 608
1933 ACAGTTACCTTGGAACTACTGATATGAGCAGGCTGCTGATCTACTAA 1982
609 ThrValThrValThrLeu.....AlaGlyGlnSerPheThr... 620
1983 GGATGACGGTGTCTACTCAAGGTATTTCACAACTTATGACACGAATGTA 2032
621 .....ThrThrValGlnAlaAspGlys 628
2033 GATACAGTGTAAAGTG.....CGGGCTCTGGAGGAGTT..... 2067
628 erTrpSerLeuThrValProAlaAlaAlaMetGlyAsnLeuProAspGly 644
2068 .....AA 2069
645 AlaValAlaIleThrAlaSerValThrAspLeuSerGlyAsnThrGlyAs 661
2070 CGAGCCAGCAGGAGAGTATACCCAGCAGAGTGGAGCAGCTGACATA. 2118
661 nThrSerArgThrIleThrValAspSerGlnAlaProAlaLeuSerIleA 678
2119 ..CCTGCTGCTGATTGAGAA.....GATGAATAACAATGGAATCCACCA 2160
678 spProLeuThrAlaAspAsnIleIleAsnAlaAlaGluSerGlyGlnAsp 694
2161 AGACCTGAATTAATAGGATGATGTTCAACACAAAGAGTGTGTTTCAG 2210
695 LeuProIleThrGlyThrThrAspAlaGlnProGlyGlnThr...Valth 710
2211 CAGNACATCTCGGAGGCTCATTTGGCTTCTGATGTCCTCCAAATGCTC 2260
710 rValThrLeuAsnGlyGlnThrThrGlnGlyValValGlnProAspGlyT 727
2261 CCATACCTGATCTCTCCACCTGGCCAAATCACCAGCTGAAGCGGAA 2310
727 hrTrpSerValThrValProAlaAlaAsnVal..... 737
2311 ATTACCGGGGCGACTCTCATT.....AATCTGACTGGACAGCTCCTGG 2354
738 .....GlyAlaLeuAlaAspGlyAsnAlaThrValThrAlaSerVa 751
2355 GGATGATTATGACCACAGCAGCTCACAAGTATATCATTCGAATAAGTA 2404
751 lAsnAspValAlaGlyAsnProSer.....Serv 761
2405 CAAGTATTTCTGTCTCAGACAGAAAGTTCATGATCTCTCAAGTGAAT 2454
761 alSerArgValAlaLeuValAspAlaThrProProValValThrIleAsn 777
2455 ACTACTGCTCTC.....ATCCCAAGGAAGCCACTCTGA 2489
778 ProValAlaThrAspAsnValIleAsnThrProGluHisAlaGlnAlaGl 794
```

2490 GGAAGTCTTTTGTTTAAACCAAGAAACATTAATTTTGAATAATGGCACAG 2539
794 nIlelle.....SerGlyThrValThrGlyAlaGlnAlaGlyA 807
2540 ATCTTTTCATTTGCTATTCAGGCTTTCATAAGCTCGATCTGAAATCAGAA 2589
807 sp.....IleValThrValThrLeuAsnValAspTyrThrThrVal 821
2590 ATATCCAAACATTCGACGAGTATCTTTGTTTATTCTCT..... 2625
822 ValAspGlySerGlyAsnTrpSerLeuGlyValProAlaSerValValSe 838
2626CCACAGACTCCCGCAGACAGACACCTAGTC 2653
838 rGlyLeuAlaAspGlySerTyrProValSerValSerValThrAspLys 855
2654 CTGATGAACGCTGCTGCTTCTGCTTAATATCATATCAACAGACCACTT 2703
855 laGlyAsnThrGlySerGlnSerLeuThrValThrValAsnThrAlaAla 871
2704 CCTGGCATTCACATTTTAAATAAT 2727
872 ProLeuIleGlyIleAsnSerIle 879
seq_name: pir2:A69222
seq_documentation_block:
hypothetical protein MTH911 - Methanobacterium thermoautotrophicum (strain Delta H)
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C;Accession: A69222
R;Smith, D.R.; Doucette-Stamm, L.A.; Delouchery, C.; Lee, H.; Dubois, J.; Aldredge, T.
; Qiu, M.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan, K.
; S. Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A;Reference number: A69000; MUID:98037514
A;Accession: A69222
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-698 <MTH>
A;Cross-references: GB:AE000866; GB:AE000666; MID:g2621998; PIDN:AAB85409.1; PID:g262
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH911
alignment_scores:
Quality: 136.00 Length: 627
Ratio: 0.437 Gaps: 28
Percent Similarity: 49.601 Percent Identity: 20.255
alignment_block:
US-09-049-696-18 x A69222 ..
Align:seg 1/1 to: A69222 from: 1 to: 698
505 CTACGATGGGAGTATT.....GA 524
1 MetLysTrpGlyPheIleLeuPheGlyMetThrLeuLeuSerGlySe 17
525 CGAGTACAATAATGATGAGAAATCTACTTATCCAAATGGAAGAAATACAG 574
17 rSerTyrGlyAlaAsp.....TyrIleValAspAsnThrThrTyrLeuG 32
575 CAGTAAAGTGTTCAGCAGGTATCTACTGTTACAAATGTTAGTAAAGAGTGT 624
32 InValPheThrProGlyGlyValGlnGlyThrPheThrLeu..... 45
625 CAGGAGGAGGCTGTTTACACCAAAAGA.....TGCACATTCATCAATAAGT 668
46 ...AsnGlyThrGluTyrThrLeuGluAspGlyAspThrPheThrPheLe 61

[illegible]

```

953  uThrLeuPheAspAlaSerAsnAlaThrArgAsnAsnLeuGluValThrL 970
2348  CTCCTCGGGATGATTGACCATGGACAGCTCACAGTAGTATCATCTCGA 2397
970  euAlaAsnGlySerValAspArgGly...AlaTrpLysTyrLysLeuArg 985
2348  ATAAGTACAAAGTATTCTTGATCTCAGA.....GACAAAGTTCAA 2435
986  AsnValAsnGlyArgTyrAspLeuTyrAsnProGluValGluLysArgAs 1002
2436  TGAATCTCTCAAGTCAAACTACTACTCTCTCATCTCCCAAGGAAGCCAACT 2485
1002  nGlnThr.....ValAspThrThr..... 1008
2486  CTGAGGAAGTCTTTTGTGTTAAACACAGAAACAATTACTTTGAAAATGGC 2535
1009  .....AsnIleThrThrProAsnAsp 1015
2536  ACAGATCTTTTTCATTCCTATTACAGCTGTGGTATGAAGTCGATCTGAATC 2585
1016  .....IleGlnAla...AspAlaProSerAlaGlnSe 1025
2585  AGAAATATCCAACTTCGACGAGTATCTTGTGTTATTCCTCCACAGACTC 2635
1025  rAsnAsnGluGluIleAlaArgValGluThrProValProProGlnAp 1042
2636  CG.....CCAGAGACACCTAGTCTCT 2655
1042  roAlaThrGluSerAlaIleAlaSerGluGlnProGluThr...ArgPro 1057
2656  GATGAACGCTGCTGCTCT 2673
1058  AlaGluThrAlaGlnPro 1063

seq_name: pir2:JC5575

seq_documentation_block:
Inter-alpha-trypsin inhibitor heavy chain 2 - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 23-Sep-1997 #sequence_revision 23-Sep-1997 #text_change
C:Accession: JC5575; PC4485
R:Nakatani, T.; Suzuki, Y.; Yamamoto, T.; Sinohara, H.
J. Biochem. 122, 71-82, 1997
A:Title: Molecular cloning and sequencing of cDNAs encoding the
sin inhibitor heavy chain family.
A:Reference number: JC5574; MUID:97420688
A:Accession: JC5575
A:Molecule type: mRNA
A:Residues: 1-946 <NAK>
A:Cross-references: DDBJ:D89286; NID:gi694689; PIDN:BAA13939.1;
A:Experimental source: liver
A:Accession: PC4485
A:Molecule type: protein
A:Residues: 55-64;140-146;151-156;424-447;500-528;577-605 <NA2>
C:Comment: In the plasma three inter-alpha-trypsin inhibitor hea
that the complexes play important role for pancreatic cancer.
C:Superfamily: inter-alpha-trypsin inhibitor complex
F:261-264,717-916/Disulfide bonds: #status predicted

```

```

alignment_scores:..
    Quality: 135.50      Length: 1008
    Ratio: 0.320         Gaps: 54
    Percent Similarity: 42.063      Percent Identity: 20.139

alignment_block:
US-09-049-696-18 x JC5575 ..

Align_seg 1/1 to: JC5575 from: 1 to: 946

      :82 CTGAGTAATTCACTCATTCAG.....CTGAACAACAATGGCTATGAAGG 125
      |||.....

```


1779 TGCTACCTGCCTCCCAATTACA.....GTGACTT 1807
553 uLeuValLeuGluThrLeuSerGlnMetAspLeuGluAspPheLeuS 570
1808 CCAAAAGCAACAGGACACACCAAAATTCGCCACCTCTGGTAGTTAT 1857
570 erLysAspLysHisAlaAspProAsnPheThrLysLysLeuTrpAlaTyr 586
1858 GCAATATTCGCCAA.....GGAGCCTCCCAATTCACG 1892
587 LeuThrIleAsnGlnLeuLeuAlaGluArgSerLeuAlaPro..... 600
1893 GGCCAGTGTACAGCCCTGATTGAATCAGTGAATGGAAACACGTT..... 1938
601ThrAlaIleLysArgLysIleThrLysThrIleLeuG 614
1939ACCTGGAACTACTG 1953
614 InMetSerLeuAspHisHisIleValThrProLeuThrAlaMetValIle 630
1954 GATAATGGAGAGGTGTGATGCTACTAAGGATGACGGTGTCTACTCAAG 2003
631 GluAsnGluAlaGlyAspGluArgMetLeuAlaAspSerPro..... 644
2004 GTATTTCACAACCTTATGACACGAATGGTAGATACAGTCTAAAGTGGCGG 2053
644 644
2054 CTCTGGGAGGAGTTAAGCGAGCCAGAGGAGTGATACCCAGCAG... 2100
645ProGlnAspHis 648
2101AGTGGAGCACTGAC..... 2115
649 SerCysSerGlyAlaLeuTyrTyrGlyThrLysValAlaLaserAlase 665
2116 ATACCTGGCTGG..... 2127
665 rIleProSerTrpAlaSerProSerProThrProValMetAlaMetLeuA 682
2128 ..ATTGAGAATGATAAATACAAATGGAATCCACCAAGACCTGAAATTAAT 2175
682 laValGlyAlaAsnArgLeuGluSerThrProProHisValIleArg 698
2176 AAGGATGATGTTCAACAC.....AAGCA 2198
699 ValGluAsnAspProHisPheIleIleTyrLeuProLysSerGlnLysAs 715
2199 AGTCTGTTTTCAGCAGAACATCTCGGGAGGCTCATTTGTGGCTTCTGATG 2248
715 nileCysPheAsnIleAspSerGlu..... 723
2249 TCCCAATGTCTCCCATACCTGATCTCTCCACCTGGCCAAATCAGCGAC 2298
724ProGlyLysIleLeuSer 729
2299 CTGAAGGGGAAATTCAGGGGGAGTCTCATTAATCTGACTGGACAGC 2348
730 LeuValSerAspProGluSerGlyIleLeuValAsn...GlyGlnLeuIl 745
2349 TCCTGGGGGATTTATGACCATGGAACAGCTCACAGTATATC..... 2391
745 eGlyAlaLysLysAlaGluAsnGlyLysLeuArgThrTyrPheGlyLysL 762
2392ATTGGAATAAGTACAAGT 2409
762 euGlyPheTyrPheGlnLysGluAspMetLysIleGluIleSerThrGlu 778
2410 ATTCTTGTATCTCAGACACAGTCTCAATGAATCTCTTCAAGTGAATACTAC 2459
779 AsnIleThrLeu.....IleAsnGlySer.....SerThrTh 789
2460 TGCTCTCATCCAAAGGAGGCC.....AACTCTGAGGAAGTCTTTT 2500

789 rSerLeuPheTrpSerAspThrAlaHisLeuGlyAsnGlnArgValLeuI 806
2501 TGTTT.....AAACCAAGAAACATTACTTTTGAATAATGGCACAGATCTT 2544
806 leSerValLysLysGlyLysSerValThrLeuThrLeuAsnLysGluMet 822
2545 TTCATTGCTATTTCAGGCTGTTGATAAGTGCATCTGAAATCAGAAATATC 2594
823 PhePheSerVal...LeuLeuHisHisValTrpLysLysHisProVal.. 837
2595 CAACATTGCAGCAGTATCTTTGTTTATTCTCCACAGACTCCGCCAGAGA 2644
838 ..AsnValAspPheLeuGlyIleTyrLeuProPro...ThrAsnLysPheS 853
2645 CACCTAGTCTCTGATGAACACGCTCTCTCTTGT.....CCTAATATT 2685
853 erProSerAlaHisGlyLeuLeuGlyGlnPheMetAsnLysProAsnIle 869
2686 CATATCAACAGCACCATTCCTGGC 2709
870 HisIlePheAsnGluArgProGly 877

This Page Blank (uspto)

About: Results were produced by the GenCore software, version 4.5, Copyright (c) 1993-2000 CompuGen Ltd.

```

seq_documentation_block:
; Sequence 4, Application US/08409995
; Patent No. 5646259
;
; GENERAL INFORMATION:
;
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Gene III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
;
; NUMBER OF SEQUENCES: 6
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hombach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA

```

ZIP: 94111-4107
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/409,995
FILING DATE: 24-MAR-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1912 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: unknown
PS-08-409-995-4

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-409-995-4 +	141.50	180.09	0.0007	1912 !
/cgn2_6/ptodata/2/iaa/6A.COMB	pep:US-08-685-467-4 +	141.50	180.09	0.0007	1912 !
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-09-377-155-33 +	141.50	177.63	0.0008	2353 !
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-08-913-947-4 +	141.50	177.63	0.0008	2353 !
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-728-470-10 +	140.50	181.16	0.0008	1529 !
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-719-641-10 +	140.50	181.16	0.0008	1529 !
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-038-682-2 +	138.50	177.96	0.0012	1536 !
/cgn2_6/ptodata/2/iaa/5A.COMB	pep:US-08-302-832-2 +	138.50	177.96	0.0012	1536 !
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-530-198-2 +	138.50	177.96	0.0012	1536 !
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-469-880-2 +	138.50	177.96	0.0012	1536 !
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-728-470-2 +	138.50	177.96	0.0012	1536 !
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-617-697-2 +	138.50	177.96	0.0012	1536 !
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-719-641-2 +	138.50	177.96	0.0012	1536 !
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-617-697-10 +	138.50	177.48	0.0012	1600 !
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-08-296-791-6 +	130.00	162.42	0.0073	1848 !
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-08-296-791-5 +	127.50	159.46	0.0116	1702 !
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-08-296-791-5 +	127.50	159.46	0.0116	1702 !
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-08-296-791-3 +	121.00	150.42	0.0408	1541 !
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-08-296-791-3 +	121.00	150.42	0.0408	1541 !
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-09-074-658-11 +	120.50	144.20	0.0572	2439 !
/cgn2_6/ptodata/2/iaa/6A.COMB	pep:US-09-045-632-28 +	118.50	157.66	0.0414	599 !
/cgn2_6/ptodata/2/iaa/6A.COMB	pep:US-09-045-632-35 +	118.50	156.84	0.0430	642 !
/cgn2_6/ptodata/2/iaa/6A.COMB	pep:US-09-045-632-25 +	118.50	153.97	0.0487	818 !
/cgn2_6/ptodata/2/iaa/6A.COMB	pep:US-09-045-632-35 +	118.50	153.97	0.0500	861 !
/cgn2_6/ptodata/2/iaa/6A.COMB	pep:US-09-045-632-21 +	118.50	152.61	0.0517	918 !
/cgn2_6/ptodata/2/iaa/6A.COMB	pep:US-09-045-632-33 +	118.50	152.07	0.0529	961 !
/cgn2_6/ptodata/2/iaa/6A.COMB	pep:US-09-045-632-16 +	118.50	151.39	0.0545	1018 !
/cgn2_6/ptodata/2/iaa/6A.COMB	pep:US-09-045-632-32 +	118.50	150.90	0.0557	1061 !
/cgn2_6/ptodata/2/iaa/6A.COMB	pep:US-09-045-632-2 +	118.50	150.34	0.0570	1112 !
/cgn2_6/ptodata/2/iaa/6A.COMB	pep:US-09-045-632-3 +	118.50	150.34	0.0570	1112 !
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-09-579-523-3 +	118.50	150.28	0.0572	1118 !
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-08-891-640-2 +	118.00	146.31	0.0727	1464 !
/cgn2_6/ptodata/2/iaa/6A.COMB	pep:US-08-604-789B-4 +	117.00	154.87	0.0571	621 !
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-09-312-721A-9 +	117.00	154.87	0.0571	621 !
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-728-470-9 +	117.00	145.80	0.0849	1338 !
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-08-719-641-9 +	117.00	145.80	0.0849	1338 !
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-09-041-886-25 +	117.00	144.87	0.0984	1447 !
/cgn2_6/ptodata/2/iaa/6B.COMB	pep:US-09-041-886-25 +	117.00	144.87	0.0984	1447 !
/cgn2_6/ptodata/2/iaa/5B.COMB	pep:US-08-617-697-9 +	117.00	143.69	0.0931	1599 !

```

alignment_scores:
  Quality: 141.50      Length: 967
  Ratio: 0.321        Gaps: 44
  Percent Similarity: 45.605      Percent Identity: 18.718

alignment_block:
  US-09-049-696-18 x US-08-409-995-4      ..
  Align seg 1/1 to: US-08-409-995-4 from: 1 to: 1912

274 ACATGGAAGACAAAGGCTGACTATGTGAGACCAAAACTTGACACTCAA 323
    ::::::::::::::::::::
428 SerTrpLysAlaLysAlaGlu..... 434
    ::::::::::::::::::::
324 AAATGCTGATGTTCTGTTGCTGAGTCTACTCTCCAGGTAATGATGAAC 373
    |||||      |||      ::::
435 ...AlaAsp.....ThrAspGlyAlaLeuGluGlyI 444
    ::::|      ::::|      ::::|
374 CCTACTACTGACGAGATGGGCAACTGTGAGGAGAAGGGTGAAGAGTCCAC 423
    ::::|      ::::|      ::::|
444 leSerLysAspGlnGluValLysAlaGlyGlu..... 454
    ::::|      ::::|      ::::|
424 CTCACCTCTGATTTTCATTGCGAGGAAAAAGATTAGCTGAATATGGACCACA 473
    |||      |||      |||
455 ...ThrValThrPheLysAlaGlyLysAsnLeu...LysValLysGlnAs 469
    ::::|      ::::|      ::::|

```

[illegible]

```

alignment_scores:
    Quality: 141.50      Length: 967
    Ratio: 0.321        Gaps: 44
    Percent Similarity: 45.605    Percent Identity: 18.718

alignment_block:
US-09-049-696-18 x US-08-685-467-4    ..

Align seg 1/1 to: US-08-685-467-4    from: 1 to: 1912

274 ACATGGAGACAAAGCGCTGACTATGTGAGACCAAACTGTGAGACCTACAA 323
    :|||:|||:|||:|||:|||:
428 SerTrpLysAlaLysAlaGlu..... 434
    ..

324 AAATGCTGATGTTCTTGGTTCCTGAGTCTACTCTCCAGGTAAATGATGAAC 373
    :|||:|||:|||:|||:
435 ....AlaAsp.....ThrAspGlyAlaLeuGluGly 444
    :|||:|||:|||:|||:
374 CCTACACTGAGCAGATGGCGAACTGTGGAGAGAGGTTGAAAGGATCCAC 423
    :|||:|||:|||:|||:
444 leSerLysAspGlnGluValLysAlaGlyGlu..... 454
    :|||:|||:|||:|||:
424 CTCACCTCCTGATTCATTCGTGAGGAAAAAGCTTACTGCAATATGACACCA 473
    :|||:|||:|||:|||:|||:
455 ...ThrValThrPheLysAlaGlyLysAsnLeu...LysValLysGlnAs 469
    :|||:|||:|||:|||:|||:

```

```

474 AGGTAGGCGATTTCACATGAG.....TGGGCTC 502
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
469 polyAlaasnPheThrTyrSerLeuGlnAspAlaLeuThrGlyLeuThrs 486
503 ATCTACGATGGGAGTATTTCACGAGTACAATAATGATGACAAATTCATC 552
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
486 erileThrLeuGlyGlyThrThrAsnGlyGlyAsnAlaLysThrVal 502
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
553 TTATCC...AATGGAAGATAACAGCAGTAAGATGTTTCAGCAGGATATAC 599
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
503 IleAsnLysAspGlyLeuThrIleThrProAlaGlyAsnGlyGlyThrTh 519
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
600 TGGTACAATGTAGTAACAGTGTACGGGAGGAGCGTGTACACCAAAA 649
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
519 rGlyThrAsnThrIleSerValThrLysAspGlyIleLysAlaGlyAsnL 536
650 GATGCACATCAATAAAGTAACAGGACATC.....TATGAAGAAAGGATGT 693
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
536 ysAlaIleThrAsnValAlaSerGlyLeuArgAlaTyrAspAlaAsn 552
694 GAGTTTGTCTCCAAATCCCGCAGCAGGAGGAGGCTTCTATAATGTTTC 743
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
553 PheAspValLeuAsnAsnSerAlaThrAsp.....LeuAs 564
744 ACAACATGTTGATCTTCTAGTTGATCTGTACAGAACAAACCAACA 793
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
564 nArgHisValGluAspAlaTyrLysGlyLeu.....AsnLeuAsnG 579
794 AAGAAGCTCCAAACAAAGCAAAATCAAAATCAATCTCCGAAGCAGATGG 843
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
579 IuLysAsnAlaAsnLysGln..... 585
844 GAAGTGATCCGTTGATCTGAGGACTTTAAGAAACCACTCTATGACAAC 893
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
586 .....ProLeuValTh 589
894 ACAGGCCAACCAATCCACTTCTCATTGCTGCAGATTGGACAAGAAATG 943
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
589 rAspSerThrAlaAlaThrValGlyAspLeu.....ArgLysLeuG 603
944 TGTGTTTGTAGCTTTCACAAATCTGGAAGCATGGCGACTGGTAACGCGCTC 993
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
603 lyTrpValValSerThrLysAsnGly.....ThrLysGluGluSer 616
994 AATCGACTGATCAACAGGAGCGAGCTTTCTCTGCTGCAGACAGTTGAGCT 1043
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
617 AsnGlnValLysGlnAlaAspGluVal.....Le 626
1044 GGGTCTGGTGGTGGATGGTGACATTTGACAGTCTGCCCATGTACAAA 1093
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
626 uPheThrGlyAlaGlyAlaAlaThrValThrSerLysSerGluAsnGlyL 643
1094 GTGAACACTCATACAGATAAACAGTGC..... 1119
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
643 ysHisThrIleThrValSerValAlaGluThrLysAlaAspCysGlyLeu 659
1120 ...AGTGACAGGACACACTCGCCAAAAGATTACCT..... 1152
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
660 GluLysAspGlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAs 676
1153 .....GCAGCGCTTCAGGAGGAGCGTCCATCTCCAGCGGCTC 1192
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
676 pAsnValLeuThrValGlyAsnAsnGlyThrAlaValThrLysGly.... 691
1193 GATCGGCATTTACTGTGATAGGAAGAATAATCCAACTGATGATCTGAA 1242
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
692 ....GlyPheGluThrVal...LysThrGlyAlaThrAspAlaAspArg 705
1243 ATTGTGCTGCTACCGATGGGGAACACACACTATAAGTGGTGTCTTAA 1292
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
706 .....GlyLysValThrValLysAspAlaThrAlaAs 716

```

```

1293 CGAGGTCAAAACAAGTGTGTCCTATCCACACAGTCGCTTTGGGGCCCT 1342
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
715 nAspAlaAspLysLysValAlaThrValLysAspValAlaThrAlaLea 733
1343 CTGACGCTCAAGAACTAGAGAGCTGCCAAATGACAGGAGGTTTACAG 1392
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
733 snSerAlaAlaThrPheValLysThrGluAsnLeuThrSerIleAsp 749
1393 ACATATGCTTCAGATCAAGTTCAGACAATGGCCTCATTTGCTTTGG 1442
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
750 .....GluAspAsnProThrAspAsnGlyLysAspAsp..... 760
1443 GGCCCTTTTCATCAGAAAT.....GGAGCTGTCTCTCAGCGCT 1480
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
761 AlaLeuLysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeu 777
1481 CCATCCAGCTTGAGAGTAAGGATTAACCTC.....CAGAACAGC 1521
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
777 ysValLysArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeu 793
1522 CAGTGGATGAATGGCAGCATGCTGGACAGCACCGTGGGAAGAGCAC 1571
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
794 GluValLysThrAlaLysValSerAspThrLeuThrIleGlyGlyAsnTh 810
1572 TTTGCTTTCTATCACCTGGACAGCGACCTCCCAATC..... 1611
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
810 r.....ProThrGlyGlyThrThrAlaThrProLysValAsnIleThrs 825
1611 ..... 1611
825 erThrAlaAspGlyLeuAsnPheAlaLysGluThrAlaAspAlaSerGly 841
1612 .....CTTCTCTGGGATCCAG 1628
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
842 SerLysAsnValTyrLeuLysGlyIleAlaThrThrLeuThrGluProse 858
1629 TGGACAAGCAAGTGGCTTGTGA.....GTGGACAAAAACACCA 1669
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
858 rAlaGlyAlaLysSerSerHisValAspLeuAsnValAspAlaThrLysL 875
1670 AATGGGCTACCTCCAAATCCAGGCATTCCTAAGTTGGCAGCTTGGAAA 1719
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
875 ysSerAsnAlaAlaSerIleGluAspValLeuArgAlaGly..... 888
1720 TACAGTCTGCAGCAAGCTCACAACTTGACCTGCTCAGCTCCCG 1769
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
889 TrpAsnIleGlnGlyAsnGlyAsnValAspTyrValAlaThrTyrAs 905
1770 TGCCTCAATGCTACC.....CTGCCTCCAATTACAGTGA 1804
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
905 pThrValAsnPheThrAspSerThrGlyThrThrValThrValT 922
1805 CTCCAAAACGACAGGACACCAGCAATTCGCCAGCCCTCTGCTAGTT 1854
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
922 hrGlnLysAlaAspGlyLysGly..... 929
1855 TATGCAATATTCGCCAGGAGCTCCCAATTCAGGGCCAGTGTAC 1904
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
920 ...AlaAspValLysIleGlyAlaLys.....Th 938
1905 AGCCCTGATTCAATCAGTGAATGAAAAACAGTTACC...TTGAACTAC 1951
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
938 rSerValIleLysAspHisAsnGlyLysLeuPheThrGlyLysAspLeuL 955
1952 TGGATAATGGAGCGTCTGATGCTACTAAGATGACGGT..... 1992
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
955 ysAspAlaAsnAsnGlyAlaThrValSerGluAspAspGlyLysAspThr 971
1993 .....GTCTACTCAAGCTATTTACAACTTATGACAGCAATGG 2030
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
972 GlyThrGlyLeuValThrAlaLys.....ThrValIleAspAlaValAs 986
2031 TAGATACAGTGTAAAAAGTGGCGGCTCTGGGA..... 2061

```

```
..... :||| :||| :|||
986 nLysSerGlyTrpArgValThrGlyGluGlyAlaThrAlaGluThrGlyA 1003
2062 .....GGAGTTACGACGACGACGACGGAGTATACCCACGACAGTGA 2106
:||||| :||||| :||||| :||||| :||||| :|||||
1003 laThrAlaValAlaAsnAlaGlyAsnAlaGluThrValThrSerGlyThrSer 1019
2107 GCACTGTACATACCTGGCTGGATTGAGATGATGAATCAATAGGAATCC 2156
: : : :|||
1020 ValAsnPhelYsnGly.....AsnAl 1027
2157 ACCAAGACCTGAATTAATAAGAT.....G 2182
:||||| :||||| :||||| :||||| :||||| :|||||
1027 aThrThrAlaThrValSerLysAspAsnGlyAsnIleAsnValLysTyra 1044
2183 ATGTT.....CAACACAGCAAGTGTGT 2205
||||| :||| :||| :||| :||| :|||
1044 spValAsnValGlyAspGlyLeuLysIleGlyAspLysLysIleVal 1060
2206 TTCAGCAGACATCC.....TCGGGAGGCTCATTTGTG..... 2238
: : :||| :||| :||| :||| :||| :|||
1061 AlaAspThrThrThrLeuThrValThrGlyGlyLysValSerValProAl 1077
2239 .....GCTTCGTGATGCTCCAAATGCTCCATACCTGATCTCTCCACCTG 2284
||||| :||| :||| :||| :||| :|||
1077 aGlyAlaAsnSerValAsnAsn..... 1084
2285 GCCAAATCACCGACTGAAGCGGGAATTCACGGGGCAGCTCTCATTAAT 2334
: : :||| :||| :||| :||| :||| :|||
1085 ..AsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsnAsn 1100
2335 CTGACTGGACACCTCTGGGGATGATTATGACCATGGACACCTCACAA 2384
||||| :||| :||| :||| :||| :|||
1101 LeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyGluSerGluGl 1117
2385 GTATATC..... 2391
1117 yGluThrAspGlnValLysAlaGlyAspLysValThrPhelLysAlaG 1134
2392 .....ATTCGAATTAAGTACAGATTCCTGATCTCAGACACAAGTTC 2433
: : :||| :||| :||| :||| :||| :|||
1134 lYlYsnLeuLysValLysGlnSerGluLysAspPheThrThrSerLeu 1150
2434 AATCAATCTCTCAAGTGAATACCTACTGCTCTCATCCCAAGGACCCAA 2483
||||| :||| :||| :||| :||| :|||
1151 GlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAlaAs 1167
2484 C.....TCGTGAGGAGCTCTTTTGTGTTTAAACCAAGAAACATTACTTTT. 2526
: : :||| :||| :||| :||| :||| :|||
1167 nGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrIle 1184
2527 .....GAAATGGCACAGAT 2541
: : :||| :||| :||| :||| :||| :|||
1184 hrLeuAlaAsnGlyAlaAlaGlyThrAspAlaSerAsnGlyAsnThr 1200
2542 CTTTTCATCTCTATCAGCTGTGTATAA..GTCGATCTCAAAATCAGAAA 2590
: : :||| :||| :||| :||| :||| :|||
1201 lIleSerValThrLysAspGlyIleSerAlaGlyAsnLysGluIleThrAs 1217
2591 TATCCAAATTCGACAGATATCTTTGTTTATCTCCACAGACTCCGCCA 2640
: : :||| :||| :||| :||| :||| :|||
1217 nValLys.....SerAlaLeuLysThrThrLysAspThrGlnAsnThrA 1232
2641 GAGACACCTAGTCTGATGAACAGCTGCTGCTTGTCTCTAATATTCATAT 2690
||| :||| :||| :||| :||| :|||
1232 laAspGlu..... 1234
2691 CAACAGCACCATCTCCGGCATTCACATTTTAAAAATATGTGGAAGTGA 2740
||| :||| :||| :||| :||| :|||
1235 ThrGlnAspLysGluPheHisAlaAlaValLysAsnAlaAsnGluValGl 1251
2741 T 2741
```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:us-09-377-155-33

seq_documentation_block:

Sequence 33, Application US/09377155

Patent No. 6197312

GENERAL INFORMATION:

APPLICANT: PEAK, Ian Richard Anselm

APPLICANT: JENNINGS, Michael Paul

APPLICANT: MOXON, E. Richard

TITLE OF INVENTION: NOVEL SURFACE ANTIGEN

FILE REFERENCE: 065064/0128

CURRENT APPLICATION NUMBER: US/09/377,155

CURRENT FILING DATE: 1999-08-19

PRIOR APPLICATION NUMBER: PCT/AU98/01031

PRIOR FILING DATE: 1998-12-14

PRIOR APPLICATION NUMBER: GB 9726398.2

PRIOR FILING DATE: 1997-12-12

NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 33

LENGTH: 2353

TYPE: PRT

ORGANISM: Haemophilus influenzae

US-09-377-155-33

alignment_scores:

Quality: 141.50 Length: 967

Ratio: 0.321 Gaps: 44

Percent Similarity: 45.605 Percent Identity: 18.718

alignment_block:

US-09-049-696-18 x US-09-377-155-33

Align seg 1/1 to: US-09-377-155-33 from: 1 to: 2353

274 ACATGGAAGCAAAAGGCTGACTATGTGAGACCAAAACTTGAGACCTACAA 323

429 SerTrpLysAlaLysAlaGlu..... 435

324 AATGCTGTGATGTTCTGTTGCTGAGTCTCTCCAGGTAATGATCAAC 373

436AlaAsp.....ThrAspGlyAlaLeuGluGlyI 445

374 CCTACACTGAGCAGATGGCACTGTGGAGAGAGGCTGAAAGGATCCAC 423

445 leSerLysAspGlnGluValLysAlaGlyGlu..... 455

424 CTCACTCCTGATTTTCATTTGAGGAAAAAGTTAGCTGAATATGACACACA 473

456 ...ThrValThrPheLysAlaGlyLysAsnLeu...LysValLysGlnAs 470

474 AGGTAGGGCATTTTCTCATCAG.....TGGGCTC 502

470 pGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuThrGlyLeuThrS 487

503 ATCTACGATGGGAGTATTTCAGCTGACATTAATGATGAGAAATCTTAC 552

487 erIleThrLeuGlyGlyThrThrAsnGlyGlyAsnAspAlaLysThrVal 503

553 TTATCC...AATGGAAGATACAGCAGTAAGATGTTTCAGCAGGTATTAC 599

504 IleAsnLysAspGlyLeuThrIleThrProAlaGlyAsnGlyGlyThrTh 520

600 TGGTACAAATCTAGTAAAGAGTGTCTAGGAGCGCAGCTGTTACACCAAAA 649

520 rGlyThrAsnThrIleSerValThrLysAspGlyIleLysAlaGlyAsnL 537

650 GATGCACATTTCAATAAAGTAACAGGACTC.....TATGAAAAAGGATGT 693

537 ysaIleThrAsnValAlaSerGlyLeuArgAlaTyrAspAlaAsn 553

694 GAGTTTGTTCCTCAATCCGCCAGCAGGAGGCTTCTATAATGTTTGC 743
|||||:|||||:|||||:|||||:LeuAs 565
554 PheAspValLeuAsnAsnSerAlaThrAsp.....LeuAs 565
744 ACAACATGTTGATCTATAGTTGAATCTGTACAGAACAAACACACA 793
:|||||:|||||:|||||:|||||: 793
565 nArgHisValGluAspAlaTyrlYsGlyLeuLeu.....AsnLeuAsnG 580
794 AGAAGCTCCAAACAAAGCAAAATCAAAATGCAATCTCCGAGCACATGG 843
:|||||:|||||:|||||:|||||: 843
580 LuLYsAsnAlaAsnLYsGln..... 586
844 GAAGTGATCCGCTGATCTGAGGACTTTAAGAAACCACTCTATGACAAC 893
:|||||:|||||:|||||:|||||:ProLeuValTh 590
587 590
894 ACAGCCACAAATCCCACTCTCTCAATGCTGCAGATTGGACAAAGATTG 943
:|||||:|||||:|||||:|||||: 943
590 rAspSerThrAlaAlaThrValIGLYsPleu.....ArgLYsLeuG 604
944 TGTGTTAGTCTTGCACAAATCTGAAGCATGGCGACTGGTAACCGCCTC 993
:|||||:|||||:|||||:|||||:ThLYsGluGluSer 617
604 lytrpValValSerThrLYsAsnGly..... 617
994 AATGACTGAATCAAGCAGCGCAGCTTTTCCGTGTCAGACAGTTGAGCT 1043
:|||||:|||||:|||||:|||||: 1043
618 AsnGlnValLYsGlnAlaAspGluVal.....Le 627
1044 GGGTCTGGTGGTGGATGGTGACATTTGCACAGTCTGCCCATGTACAAA 1093
:|||||:|||||:|||||:|||||: 1093
644 ysHisThrIleThrValSerValAlaGluThrLYsAlaAspCysGlyLeu 660
1120 ...AGTGACAGGACACACTCCGCCAAAGATTACCT..... 1152
661 GluLYsAspGlyAspThrIleLYsLeuLYsValAspAsnGlnAsnThrAs 677
1153CGACGAGCTTCAGGAGGAGTCCATCTGCACGGGCTC 1192
677 pAsnValLeuThrValIGLYsAsnGlyThrAlaValThrLYsGly.... 692
1193 GATCGGCACTTACTGTGATAGGAAGAATATCCAACGTGATGACTGAA 1242
:|||||:|||||:|||||:|||||: 1242
693GlyPheGluThrVal...LYsThrGlyAlaThrAspAlaAspArg 706
1243 ATGTGCTCTGACCGATGGGAGACACACTATAAGTGGTGCTTTAA 1292
:|||||:|||||:|||||:|||||: 1292
707GlyLYsValThrValLYsAspAlaThrAlaAs 717
1293 CGAGGTCAACAAAGTGGTGCATCATCCACAGCTCGCTTTGGGGCCT 1342
:|||||:|||||:|||||:|||||: 1342
717 nAspAlaAspLYsLYsValAlaThrValLYsAspValAlaThrAlaIleA 734
1343 CTGCAGCTCAAGACTAGAGGAGCTGTCCAAATGACAGGAGTTTACAG 1392
:|||||:|||||:|||||:|||||: 1392
734 snSerAlaAlaThrPheValLYsThrGluAsnLeuThrThrSerIleAsp 750
1393 ACATATGCTTCAGATCAAGTTCAGAACAAATGGCTCATTTGCTTTGG 1442
:|||||:|||||:|||||:|||||: 1442
751GluAspAsnProThrAspAsnGlyLYsAspAsp..... 761
1443 GGCCCTTTTCATCAGAAAT.....GGAGCTGTCTCTCAGCGCT 1480
:|||||:|||||:|||||:|||||: 1480
762 .AlaLeuLYsAlaGlyAspThrLeuThrPheLYsAlaGlyLYsAsnLeuL 778
1481 CCATCCAGCTTGAGTAGGATTAAACCTC.....CAGAACACG 1521
:|||||:|||||:|||||:|||||: 1521
778 ysValLYsArgAspGlyLYsAsnIleThrPheAspLeuAlaLYsAsnLeu 794

1522 CAGTGGATGAATGCCACAGTGCATCGTGGACAGCAGCCGTGGAAAGGACAC 1571
:|||||:|||||:|||||:|||||: 1571
795 GluValLYsThrAlaLYsValSerAspThrLeuThrIleGlyGlyAsnTh 811
1572 TTTGTTTCTTATCACCTGGACAGCAGCCCTCCCAATC..... 1611
811 r.....ProThrGlyGlyThrThrAlaThrProLYsValAsnIleThrS 826
1611 1611
826 erThrAlaAspGlyLeuAsnPheAlaLYsGluThrAlaAspAlaSerGly 842
1612CTTCTCTGGGATCCAG 1628
843 SerLYsAsnValTyrlYsGlyIleAlaThrThrLeuThrGluProse 859
1629 TGGACAGACAGCAGTGGCTTGTGA.....GTGACAAAACACCA 1669
:|||||:|||||:|||||:|||||: 1669
859 rAlaGlyAlaLYsSerSerHisValAspLeuAsnValAspAlaThrLYsL 876
1670 AAATGGCTTACCTCCAATCCAGGCATTGCTAAGTTTGGCACTTTGAAA 1719
:|||||:|||||:|||||:|||||: 1719
876 ysSerAsnAlaAlaSerIleGluAspValLeuArgAlaGly..... 889
1720 TACAGTCTGCAAGCAAGCTCACAACTTGACCTGACTGTCACGTCCCG 1769
:|||||:|||||:|||||:|||||: 1769
890 TrpAsnIleGlnGlyAsnGlyAsnValAspTyrlValAlaThrTyraS 906
1770 TGCCTCCAATGCTACC.....CTGCCTCCAATTTACAGTGA 1804
:|||||:|||||:|||||:|||||: 1804
906 pThrValAsnPheThrAspSerThrGlyThrThrValThrValt 923
1805 CTTCAAAACGACAAAGACACACCACTGACCAATTCGCCAGCCCTCTGTTAGT 1854
:|||||:|||||:|||||:|||||: 1854
923 hrGlnLYsAlaAspGlyLYsGly..... 930
1855 TATCAATATTCGCCAAGGAGCTCCCAATTCAGGGCCAGTGTAC 1904
931 ...AlaAspValLYsIleGlyAlaLYs.....Th 939
1905 AGCCTGATTTGAATCAGTGAATGAAAACAGTTACC...TTGAACTAC 1951
:|||||:|||||:|||||:|||||: 1951
939 rSerValIleLYsAspHisAsnGlyLYsLeuPheThrGlyLYsAspLeuL 956
1952 TGGATATGGACAGGCTGCTGCTACTACTAAGGATGACGGT..... 1992
:|||||:|||||:|||||:|||||: 1992
956 ysAspAlaAsnAsnGlyAlaThrValSerGluAspAspGlyLYsAspThr 972
1993GTCTACTCAAGGTATTTCAAACTTATGACACGAATGG 2030
:|||||:|||||:|||||:|||||: 2030
973 GlyThrGlyLeuValThrAlaLYs.....ThrValIleAspAlaValAs 987
2031 TAGATACAGTGTAAAGTGGCGGCTCTGGGA..... 2061
:|||||:|||||:|||||:|||||: 2061
987 nLYsSerGlyTrpArgValThrGlyGlyAlaThrAlaGluThrGlyA 1004
2062GGAGTTAAAGCAGCAGAGAGAGATACCCAGCAGAGTGA 2106
1004 laThrAlaValAsnAlaGlyAsnAlaGluThrValThrSerGlyThrSer 1020
2107 GCAGTGTATACCTGGCTGGATTGAGAATGATGAATACAAATGGAATCC 2156
:|||||:|||||:|||||:|||||: 2156
1021 ValAsnPheLYsAsnGly.....AsnAl 1028
2157 ACCAAGACTGAATTAATTAAGGAT.....G 2182
:|||||:|||||:|||||:|||||: 2182
1028 aThrThrAlaThrValSerLYsAspAsnGlyAsnIleAsnValLYsTyra 1045
2183 ATGTT.....CAACACAAGCAAGTGTGT 2205
1045 spValAsnValGlyAspGlyLeuLYsIleGlyAspAspLYsLYsIleVal 1061
2205 TTCAGCAGAACATCC.....TCGGAGGCTCATTTGTG..... 2238

```
... |||... :|||...
1062 AlaAspThrThrLeuThrValThrGlyGlyLysValSerValProAl 1078
2239 ...GCTTCTGATGTCCTCCAAATGCTCCCATACCTGATCTCTTCCACCTG 2284
|||...:|||||
1078 aGlyAlaAsnSerValAsnAsn..... 1085
2285 GCCAATACCCGACCTGAAGGGGGAATTCACGGGGGCGAGTCTCATTAAT 2334
|||...:||||| :| :| :| :| :|
1086 ..AsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsnAsn 1101
2335 CTGACTTGACAGCTCCTCGGGATGATTATGACCATGGAACAGCTCACAA 2384
|||...:||||| :| :| :| :| :|
1102 LeuSerTrpThrAlaLysAlaLysLysTyrAlaAspGlyLysSerGluG 1118
2385 GTATATC..... 2391
1118 yGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysAlaG 1135
2392 .....ATTCGAATAAGTACAAAGTATTCTTGATCTCAGACAAAGTTC 2433
|||...:||||| :| :| :| :| :|
1135 LysAsnLeuLysValLysGlnSerGlnLysAspPheThrTyrSerLeu 1151
2434 AATGAATCTCTCAAGTGAATACTACTGCTCTCATCCCAAGGAAGCCAA 2483
|||...:||||| :| :| :| :| :|
1152 GlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAlaAs 1168
2484 C.....TCTGAGGAAGTCTTTTGTGTTTAAACACAGAAACATTACTTTT. 2526
|||...:||||| :| :| :| :| :|
1168 nGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrIleT 1185
2527 .....GAAATGGCACAGAT 2541
|||...:|||||
1185 hrLeuAlaAsnGlyAlaAlaGlyThrAspAlaSerAsnGlyAsnThr 1201
2542 CTTTTCATTGCTATTGAGGCTGTGATAA. GGTGATCTGAAATCAGAAA 2590
|||...:||||| :| :| :| :| :|
1202 IleSerValThrLysAspGlyIleSerAlaGlyAsnLysGluLeuThrAs 1218
2591 TATCCAACTTGCACAGTATCTTTTCTTTTATCTCTCCACAGACTCGGCA 2640
|||...:||||| :| :| :| :| :|
1218 nValLys.....SerAlaLeuLysThrLysAspThrGlnAsnThrA 1233
2641 GAGACACCTAGTCTGATGAAAGCTGCTCTCTTGTCTCTTAATTCATAT 2690
|||
1233 IaAspGlu..... 1235
2691 CAACAGCACCATTCCTGGCATTCACATTTTAAATAATTATGTGGAAGTGA 2740
|||...:||||| :| :| :| :| :|
1236 ThrGlnAspLysGluPheHisAlaAlaValLysAsnAlaAsnGluValG 1252
2741 T 2741
1252 u 1252
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:us-08-913-942-4

```
seq_documentation_block:
; Sequence 4, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hobbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
```

```
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,942
FILING DATE: 29-DEC-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/4031
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vance, Dolly A.
REGISTRATION NUMBER: 39,054
REFERENCE/DOCKET NUMBER: A-61053-1/RET/RMS/DAV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2353 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-913-942-4

alignment_scores:
Quality: 141.50 Length: 967
Ratio: 0.321 Gaps: 44
Percent Similarity: 45.605 Percent Identity: 18.718

alignment_block:
US-09-049-696-18 x US-08-913-942-4 ..
Align seg 1/1 to: US-08-913-942-4 from: 1 to: 2353
274 ACATGGAAGACAAAGGCTGACATATGTGACCAACCAAACTTGAGACCTACAA 323
429 SerTrpLysAlaLysAlaGlu..... 435
324 AAATGCTGATGTTCTGTTGCTGCTACTCTCCAGGTAAATGATGAAC 373
|||...:||||| ThrAspGlyAlaLeuGluGlyI 445
436 ....AlaAsp..... 445
374 CTTACTGAGCAGATGGGCACTGTGGAGAGAAGGTGAAGGATCCAC 423
|||...:|||||
445 leSerLysAspGlnGluValLysAlaGlyGlu..... 455
424 CTCCTCTGATTTTCATTCAGGAAAAAGTTAGCTGAATATGACCAACA 473
|||...:||||| :| :| :| :| :|
456 ...ThrValThrPheLysAlaGlyLysAsnLeu...LysValLysGlnAs 470
474 AGGTAGGCGATTGTCTCATCAG.....TGGGCTC 502
|||...:||||| :| :| :| :| :|
470 pGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuThrGlyLeuThrS 487
503 ATCTAGATGGGAGTATTTGACGAGTACAATAATGATGAGAAATTCAC 552
|||...:||||| :| :| :| :| :|
487 erIleThrLeuGlyGlyThrThrAsnGlyLysAsnAspAlaLysThrVal 503
553 TATATCC....AATGGAAGAATACAAAGCAAGTGTTCACAGGATTTAC 599
|||...:||||| :| :| :| :| :|
504 IleAsnLysAspGlyLeuThrIleThrProAlaGlyAsnGlyGlyThrTh 520
600 TGGTACAAATGTAGTAAAGAAGTGTCCAGGAGGCGACGCTGTTCACCAAAA 649
|||...:||||| :| :| :| :| :|
520 rGlyThrAsnThrIleSerValThrLysAspGlyIleLysAlaGlyAsnL 537
```

650 GATGCACATTCATTAAGATCAACAGACTC.....TATGAAAAAGGATGT 693
 :::: ||| :::::||||| ||::: :::
 537 ysaIaIleThrAsnValAlaSerGlyLeuArgAlaTyAspAspAlaAsn 553
 694 GAGTTGTGTTCCATCCCGCCAGACGCGAGAGAGCTTCTATATGTTTGC 743
 :|||||::: ||::: ||:::
 554 pheAspValLeuAsnAsnSerAlaThrAsp.....LeuAs 565
 744 ACAACATGTTGATTCATATAGTTGAATTCGTACAGAACAAAAACCACAACA 793
 ::::|||||::: ||| ||::: ||:::
 565 nargHisValGluAspAlaTyLysGlyLeu.....AsnLeuAsn 580
 794 AAGAAGCTCCAAACAGCAAAATCAAAATGCAATCTCCGAAGCAGCATGG 843
 :::: ||||||| ||:::
 580 luLysAsnAlaAsnLysGln..... 586
 844 GAAGTGATCCGTGATCTTGAGGACTTTAAGAAAAACCCTCTATGACAAC 893
 :|||||::: ||::: ||::: ||::: ||::: ||:::
 587ProLeuValTh 590
 894 ACAGCCACCAAAATCCACCTTCTCATTCGTCTGAGATGGACAAGAATTG 943
 :::: ||| ::: ||| ||::: ||:::
 590 rAspSerThrAlaAlaThrValGlyAspLeu.....ArgLysLeuG 604
 944 TGTGTTAGTCCTGACAAATCTGGAAGCATGGCGACTGTAAACCGCTC 993
 :||| |||::: ||| ||::: ||:::
 604 lYtrpValValSerThrLysAsnGly.....ThrLysGluGluSer 617
 994 AATCGACTCAATCAGACGGCCAGCTTTTCTGCTGCTGACGACAGTTGAGCT 1043
 :|||||::: ||::: ||::: ||::: ||::: ||:::
 618 AsnGlnValLysGlnAlaAspGluVal.....Le 627
 1044 GGGGCTCTGGTGGGTGGATCTTGACATTTGACAGTGTGCCCATGTACAAA 1093
 :||| |||::: ||| ||::: ||:::
 627 uPheThrGlyAlaGlyAlaAlaThrValThrSerLysSerGluAsnGlyL 644
 1094 GTGAACATCATAGATAAACAGTGGC..... 1119
 :||| |||::: ||: ||::: ||:::
 644 ySHisThrIleThrValSerValAlaGluThrLysAlaAspCysGlyLeu 660
 1120 ...AGTGACGGGACACACTCGCCAAAAGATTACT..... 1152
 :|||| |::: ||::: ||::: ||:::
 661 GluLysAspGlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAs 677
 1153CGACAGCTTCAGGAGGAGCTCCATCTGACAGCGGGCTC 1192
 :::: |||::: ||| ||::: ||::: ||:::
 677 pAsnValLeuThrValGlyAsnAsnGlyThrAlaValThrLysGly... 692
 1193 GATCGGCATTTACTGTGATTAGGAAGAATAATCAACTGATGATCTCAA 1242
 :||| |||::: ||| ||::: ||::: ||:::
 693GlyPheGluThrVal...LysThrGlyAlaThrAspAlaAspArg 706
 1243 ATTGTGCTGCTACGGATGGGAGACAACTATAGTGGGTCTTTAA 1292
 :|||::: ||::: ||::: ||::: ||:::
 707GlyLysValThrValLysAspAlaThrAlaAs 717
 1293 CGAGGTCAAAACAAAGTGTGGCATCATCCACAGCTCGCTTGGGGCCCT 1342
 :|||::: ||| ||::: ||::: ||::: ||:::
 717 nAspAlaAspLysLysValAlaThrValLysAspValAlaThrAlaIle 734
 1343 CTGCGACTCAAGAACTAGAGGAGCTGTCCAAATATCAGAGAGGTTACAG 1392
 :||| |||::: ||| ||::: ||::: ||:::
 734 snSerAlaAlaThrPheValLysThrGluAsnLeuThrThrSerIleAsp 750
 1393 ACATATGCTTCAGATCAGTTGACAAACATGGCTCATATGATGCTTTGG 1442
 :|||::: ||::: ||::: ||::: ||:::
 751GluAspAsnProThrAspAsnGlyLysAspAsp..... 761
 1443 GGCCCTTCATCAGGAAT.....GGAGCTGTCTCTCAGCGCT 1480
 :|||||::: ||| ||::: ||::: ||:::
 762 .AlaLeuLysAlaGlyAspThrLeuThrPheLysAlaGlyLysLeuL 778
 1481 CCATCCAGCTTCAGAGTAAGGATTAAACCTTC.....CAGAACAGC 1521

778 ysVallysArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeu 794
 1522 CAGTGGATGAATGGCACAGTCATCGTGCAGCAGCACCTGGGAAAGACAC 1571
 795 GluValLysThrAlaLysValSerAspThrLeuThrIleGlyGlyAsnTh 811
 1572 TTGTGTTCTTATCACCTGCAGCAAGCAGCCCTCCCAATC..... 1611
 811 r.....ProThrGlyGlyThrThrAlaThrProLysValAsnIleThrS 826
 1611 1611
 826 exThrAlaAspGlyLeuAsnPheAlaLysGluThrAlaAspAlaSerGly 842
 1612CTTCTCTGGGATCCCG 1628
 843 SerLysAsnValTyLeuLysGlyIleAlaThrThrLeuThrGluProSe 859
 1629 TGCACAGAAGCAAGTGGCTTTGTA.....GTGGCAAAAACACCA 1669
 859 rAlaGlyAlaLysSerSerHisValAspLeuAsnValAspAlaThrLysL 876
 1670 AAATGGCCTACCTCCAATCCAGCATGTCTAGTTGGACTTGGGAA 1719
 876 ysSerAsnAlaAlaSerIleGluAspValLeuAlaGly..... 889
 1720 TACAGTCTCAAGCAGCTCAACAACCTTGACCTGACTGTCAGTCCCG 1769
 890 TPasnIleGlnGlyAsnGlyAsnValAspTyrValAlaThrTyrAs 906
 1770 TGCCTCCAATGCTACC.....CTGCCCTCCAATTACAGTGA 1804
 906 pThrValAsnPheThrAspAspSerThrGlyThrThrValThrValT 923
 1805 CTTTCCAAAACGAAGCAGCACCCAGCAAAATTTCCCGACCTCTGGTAGTT 1854
 923 hrClnLysAlaAspGlyLysGly..... 930
 1855 TATGCAAAATTTCCGAAGGACCTCCCAATTTCTAGGCCAGTGTCA 1904
 931 ...AlaAspValLysIleGlyAlaLys.....Th 939
 1905 AGCCCTGATTGAATCAGTGAATGGAAAAACAGTTACC...TTGGAACTAC 1951
 939 rSerValLleLysAspHisAsnGlyLysLeuPheThrGlyLysAspLeuL 956
 1952 TGGATAATGGACAGGTGCTGATGCTACTAAGGATGACGGT..... 1992
 956 ysAspAlaAsnAsnGlyAlaThrValSerGluAspAspGlyLysAspThr 972
 1993GTCTACTCAAGTATTTCACACTTATGCACGAGATGG 2030
 973 GlyThrGlyLeuValThrAlaLys.....ThrValIleAspAlaValas 987
 2031 TAGATACAGTGAAGAGTCGGGCTCTCGGA..... 2061
 987 nLysSerGlyTyrArgValThrGlyGluGlyAlaThrAlaGluThrClyA 1004
 2062GGAGTTAACCGCACGAGCGAGCTGATACCCACGACAGTGG 2106
 1004 lathrAlaValAsnAlaGlyAsnAlaGluThrValThrSerGlyThrSer 1020
 2107 GCAGTGTACATACCTGGCTGGATTGAGAAATACAAATGGAATCC 2156
 1021 ValAsnPheLysAsnGly.....AsnAl 1028
 2157 ACCAGACCTGAATTAATTAAGTAT.....G 2182
 1028 aThrThrAlaThrValSerLysAspAsnGlyAsnIleAsnValLysTyrA 1045
 2183 ATGTT.....CAACACAGCAAGTGTGT 2205
 1111

STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728.470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-10

alignment_scores:		
Quality:	140.50	Length: 680
Ratio:	0.431	Gaps: 29
Percent Similarity:	47.941	Percent Identity: 20.882

seq_documentation_block:
; Sequence 10, Application US/08728470
; Patent No. 5928651

1033ACAGTTGACTGGGGPCCCTGGGTGGATGGTGACATTTCAC... 107
60 nGlyIleThrIIeGlyLysAspAlaIleIleAsnThrAsnGlyPheThrA 97
||||| : : : : : ||||
1075 ..AGTGTCTCCCATTGCACAAAGTGAACTCATACAGATAAACAGTCGCAGT 1122
||||| : : : : : ||||| : : : : : CAGT
97 laSerThrLeuAspIleSerAsnGluAsnIleLysAlaArgAsnPheThr 113
||||| : : : : : ||||| : : : : :
1123 GACAGGGACACACTCGCCCAAAAGATTTCCTGTGCAGCAGCTTCAGGAGGGAC 1172

114 LeuGluGlnThrLysAspLysAlaLeuAlaGluIleValAsnHisGlyLe 130
1173 CTCATCTGCACGGCGCTTCGATCGCATTTACTGTGATTAGGAAGAAT 1222
130 u...ilethrValGlyLysAspGlySerValAsnLeuileGlyGlyLysV 146
1223 ATCCAACTGAT.....GGATCTGAATTTGCTGCTG 1254
146 alLysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeuLeu 162
1255 ACGGATGGGAAGACACACATATAGTGGTGCTTTACAGAGGTCARACA 1304
163 Ala...GlyGlnIleThrIleSerAspIleIleAsnPro..... 175
1305 AAGTGGTCCCATCTACACAGCTGCTTTGGGGCCCTCTCCAGCTCAAG 1354
176ThrIleThrTyrSerIleAlaAlaProGluAsnGluAlaIleA 190
1355 AACTAGGAGCTGTCCAAATACAGAGGTTTACAGACATATGCTTCA 1404
190 snLeuGlyAspIlePheAlaLysGlyGlyAsnIleAsnValArgAlaAla 206
1405 GATCAAGTTCAGAACATGGCTCAT.....GATGCTTTTGGGGCCCT 1448
207 ...ThrIleArgAsnLysGlyLysLeuSerAlaAspSerValSerLysAs 222
1449 TTCATCAGAAATGAGCTGCTCTCAGCGCTCCATCCAGCTTCAGAGTA 1498
222 pLysSerGlyAsnIleValLeuSerAlaLysGluGlyGluAlaGluIleG 239
1499 AGGA...TTAACCTCCAGAACAGCCAGTGGATGATGGACAGCTGATC 1545
239 lyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGlyLysLeuMet 255
1546 GTG..... 1548
256 ileThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLeuSe 272
1549GACAGCACCGGTG 1561
272 rGlyLysGluGlyGlyGluThrTyrLeuGlyLysGluArgGlyGluG 289
1562 GAAGGACACTTTGTTCTTATCCTGGACAGCCGCTCCCAATC 1611
289 lyLysAsnGlyIleGlnLeuAlaLysTyrThrLeuGluLysGlySer 305
1612 CTCTCTGGGATCCCGTGGACAGCAAGGTGGC..... 1647
306 ThrIle...AsnValSerGlyLysGluLysGlyArgAlaIleValTr 321
1648TTGTAGTGGACAAAACACCAAAATGGCTTACCTCCAAA 1687
321 pGlyAspIleAlaLeuIleAspGlyAsn.....AlaGlnG 335
1688 TCCAGGATTCGTAAGTGGCACTTGGAAATACAGTCTCAAGCAAGC 1737
335 lySerAspIleAlaLysThrGlyGlyPhe.....ValGluThrSer 348
1738 TCACAAACCTTGACCTG..... 1755
349 GlyHisAspLeuSerIleGlyAspValIleValAlaLysGluTr 365
1756ACTGTCACGTCCTCGCTGCT 1774
365 pLeuLeuAspProAspAspValSerIleGluThrLeuThrSerGlyArgA 382
1775 CCAATGTCACCTGCTCCATTCAGATGACTTCCAAACGACAGGAC 1824
382 snAsnThrGlyGluAsnGlnGlyTyrThrThrGlyAspGlyThrLysGlu 398
1825 ACCAGCAAA.....TTCCCGACGCTCTGGTAGTTTATGCAAAATAT 1865
399 SerProLysGlyAsnSerIleSerLysProThrLeuThrAsnSerThrLe 415

1866 TCGCCAAAGAGCCTCCCAATTCACAGGCGCAGTGTACAGCCCTGATTG 1915
415 uGluGln.....IleLeuArgArgGlySerTyrValAsnIleT 428
1916 AATCAGTGAATGGAAACACAGTTACCTTGGAACTA...CTGGATAATGGA 1962
428 hrAlaAsnAsnArgIleTyrValAsnSerIleAsnLeuSerAsnGly 444
1963 CGAGGTGCTGATGCTACTAAGGATGACGCTGTCTACTCAAGGTATTTTCC 2012
445 SerLeuThrLeuHisThrLysArgAspGlyValLysIleAsnGlyAspI 461
2013 AACTTATGACACGAATGTAGTACAGTGTAAAGTGGGCTCTGGAG 2062
461 eThrSerAsnGluAsnGlyAsnLeuThrIleLysAlaGlySerTrpVala 478
2063 GAGTTAAGCCAGCAGAGGAGAGTATACCCAGCAGAGTGGAGCACTG 2112
478 spValHis.....LysAsnIleThrLeuGlyThrGlyPheLeu 490
2113 TACATACCTGGCTGGATTGAGAATGATAAATACAATGGAATCCACCAAG 2162
491 AsnIle.....ValAlaGlyAspSerValAlaPheGluArgGluG 504
2163 ACCTGAA..... 2169
504 yAspLysAlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrI 521
2170ATTATAAGATGATGTTCAACACAAG.....CAAGTGTGTTTC 2208
521 leThrValAsnLysAspLysGlnPheArgPheAsnAsnValSerIle 537
2209 AGCAGACATCCTCGGAGGCTCATTTGTGCTTCTGATGTCCTCCAAATGC 2258
538 AsnGlyThrGlyLysGlyLeuLysPheIleAlaAsnGln..... 550
2259 TCCATACCTGATCTCTTCCACCTGGCCAAATCACC.....GACC 2299
551AsnAsnPheThrHisLysPheAspG 559
2300 TGAAGCGGAAATTCACGGGGCAGTCTCATTAATCTGACT..... 2340
559 lyGluLeuAsnIleSerGlyIleValThrIleAsnGlnThrThrLysLys 575
2341TGCACAGCTCTCGGATGATGATGACCAT..... 2370
576 AspValLysTyrTrpAsnAlaSerLysAspSerTyrTrpAsnValSerSe 592
2371GGACAGCTCACAGTAT.....ATCATTCGAA 2398
592 rLeuThrLeuAsnThrValGlnLysPheThrPheIleLysPheValAsps 609
2399 TAAGTACAAATGATTTCTGATCTCAGACACAAGTTCAATGAATCTCTCAA 2448
609 erGlySerAsnSerGlnAspLeuArgSerSerArgArgSerPheAlaGly 625
2449 CTGAATACTACTGCTCTCATCCCAAGGAA.....GCCAA 2483
626 ValHisPheAsnGlyIleGlyLysThrAsnPheAsnIleGlyAlaAs 642
2484 CTCGAGGAAGTCTTTTGTGTTTAAACAGAAACATTTACT 2523
642 nAlaLysAlaLeuPheLysLeuLysProAsnAlaAlaThr 655

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep-US-08-719-641-10

seq_documentation_block:

; Sequence 10, Application US/08719641

; Patent No. 6218141

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins

```

alignment_scores:
  Quality: 140.50      Length: 680
  Ratio: 0.431        Gaps: 29
  Percent Similarity: 47.941  Percent Identity: 20.882

alignment_block:
  US-09-049-696-18 x US-08-719-641-10  ..

  Align seg 1/1 to: US-08-719-641-10 from: 1 to: 1529

865  GACTTTAAAGAAACCACCTCTATGACAACACAGCCACCACCAATCCACCTT 914
    ||| ::||| ||||| ::||| ::|||
14  AspGlyAsnLysThrThrIleArgAsnSerValAsnAlaIleIleAsnTr 30

915  CTCATGTGTCAGATGGACAAGAANTTGTGTTTAGTCCTTGCACAAAT 964
    ::||| ::||| ::||| ::||| ::||| ::|||
30  pLysGlnPheAsnIleAspGlnAsnGluMetGluGlnPheLeuGlnGluS 47

965  CTGGAAGCATGCGGACTGTGTATACCGCCTCAATCGA..... 999
    ||| ::||| ::||| ::||| ::||| ::|||
47  erSerAsnSerAlaValPheAsnArgValThrSerAspGlnIleSerGln 63

1000 .....CTGAATCAACGAGCCAGCTTTTCTCGTCGAG..... 1032
    ||| ::||| ::||| ::||| ::||| ::|||
64  LeuLysGlyIleLeuAspSerAsnGlyGlnValPheLeuIleAsnProAs 80

1033 .....ACAGTTAGCTGGGCTCTGGGTTGGGATGGTGACATTTGAC. 1074
    ||| ::||| ::||| ::||| ::||| ::|||
80  nGlyIleThrIleGlyLysAspAlaIleIleAsnThrAsnGlyPheThrA 97

```

```
1775 CCAATGCTACCCCTGCTCCATTCAGTACAGTCTCCAAACGACGAC 1824
      ::::: ::::: ::::: :::::
382 sAsnThrGlyLeuAsnGlnGlyThrThrGlyAspGlyThrLysGlu 398
      ::::: ::::: ::::: :::::
1825 ACCAGCAA.....TCCCAGCCCTCTGGTAGTTTATGCAAAAT 1865
      ::::: ::::: ::::: :::::
399 SerProLysGlyAsnSerIleSerLysProThrLeuThrAsnSerThrLe 415
      ::::: ::::: ::::: :::::
1866 TCGCAAGAGCCCTCCCAATCTCAGGCCAGTGTCACAGCCCTGATG 1915
      ::::: ::::: ::::: :::::
415 uGluGln.....IleLeuArgArgGlySerTyrValAsnIleT 428
      ::::: ::::: ::::: :::::
1916 AATCAGTCAATGGAAAAACAGATTACCTTGGAACTA...CTGCAATATGA 1962
      ::::: ::::: ::::: :::::
428 hrAlaAsnAsnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGly 444
      ::::: ::::: ::::: :::::
1963 GCAGTGCTGATGCTACTAAGATGACGGGTGCTACTCAAGGTATTTCAC 2012
      ::::: ::::: ::::: :::::
445 SerLeuThrLeuHisThrLysArgAspGlyValLysIleAsnGlyAspI 461
      ::::: ::::: ::::: :::::
2013 AACTTATGACAGCAATGCTAGTACAGTGTAAAGTGGGCTCTGGAG 2062
      ::::: ::::: ::::: :::::
461 eThrSerAsnGluAsnGlyAsnLeuThrIleLysAlaGlySerTrpValA 478
      ::::: ::::: ::::: :::::
2063 GAGTTAACCCAGCCAGACGGAGAGTGATACCCACAGACAGTGGACCTG 2112
      ::::: ::::: ::::: :::::
478 spValHis.....LysAsnIleThrLeuGlyThrGlyPheLeu 490
      ::::: ::::: ::::: :::::
2113 TACATACCTGGCTGATGAGTGAATGATGAATACAAATGGAATCCACCAAG 2162
      ::::: ::::: ::::: :::::
491 AsnIle.....ValAlaGlyAspSerValAlaPheGluArgGluG 504
      ::::: ::::: ::::: :::::
2163 ACCTGAA..... 2169
      ::::: ::::: ::::: :::::
504 yAspLysAlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrI 521
      ::::: ::::: ::::: :::::
2170 .....ATTAATAAGGATGATCTCAACACAAG.....CAAGTGTGTTTC 2208
      ::::: ::::: ::::: :::::
521 leThrValAsnLysAspLysGlnPheArgPheAsnValSerIle 537
      ::::: ::::: ::::: :::::
2209 AGCAAGACATCCTCGGAGGCTCAATTTCTGCTTCTGATGCCCAATGC 2258
      ::::: ::::: ::::: :::::
538 AsnGlyThrGlyLysGlyLeuLysPheIleAlaAsnGln..... 550
      ::::: ::::: ::::: :::::
2259 TCCCATACCTGATCTCTCCACCTGGCCAAATCACC.....GACC 2299
      ::::: ::::: ::::: :::::
551 .....AsnAsnPheThrHisLysPheAspG 559
      ::::: ::::: ::::: :::::
2300 TCAAGCGGAAATTCACGGGGCAGCTCTCATTAATCTGACT..... 2340
      ::::: ::::: ::::: :::::
559 yGluLeuAsnIleSerGlyIleValThrIleAsnGlnThrThrLysLys 575
      ::::: ::::: ::::: :::::
2341 .....TGACAGCTCTGGGATGATTATGACCAT..... 2370
      ::::: ::::: ::::: :::::
576 AspValLysTyrTrpAsnAlaSerLysAspSerTyrTrpAsnValSerSe 592
      ::::: ::::: ::::: :::::
2371 .....GGAACAGCTCAACAGTAT.....ATCATTCGAA 2398
      ::::: ::::: ::::: :::::
592 rLeuThrLeuAsnThrValGlnLysPheThrPheIleLysPheValAsps 609
      ::::: ::::: ::::: :::::
2399 TAAGTACAAGTATTCTTGATCTCAGACAGCAAGTTCAAATCTCTTCAA 2448
      ::::: ::::: ::::: :::::
609 erGlySerAsnSerGlnAspLeuArgSerSerArgSerPheAlaGly 625
      ::::: ::::: ::::: :::::
2449 GTGATACTACTGCTCTCATCCCAAGAA.....GCCAA 2483
      ::::: ::::: ::::: :::::
626 ValHisPheAsnGlyIleGlyGlyThrAsnPheAsnIleGlyAlaAs 642
      ::::: ::::: ::::: :::::
2484 CTCTGAGGAAGTCTTTTGTAAACCAAGAAACATTACT 2523
      ::::: ::::: ::::: :::::
642 nAlaLysAlaLeuPheLysLeuLysProAsnAlaAlaThr 655
      ::::: ::::: ::::: :::::
seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.US-08-038-682-2
```

```
seq_documentation_block:
Sequence 2, Application US/08038682
Patent No. 5549897
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038,682
FILING DATE: 16-MAR-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-038-682-2

alignment_scores:
Quality: 138.50 Length: 936
Ratio: 0.318 Gaps: 49
Percent Similarity: 46.581 Percent Identity: 20.406

alignment_block:
US-09-049-696-18 x US-08-038-682-2 ..
Align seg 1/1 to: US-08-038-682-2 from: 1 to: 1536
217 TTTGAAGCTACA.....GGAAAGCGATTTTATTTCAAAAATGT 254
      ::::: :::::
631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnI 643
255 TGCATTTTGATTCTCGAAACATGGAACACAGCTGACTATGTGAGAC 304
      ::::: :::::
643 eSerMetValLeuPro.....LysAsnGluSerGlyTyrAspLysp 657
      ::::: :::::
305 CAAAACTTGAGACCTACAAAATGCTGATGTTCTG...GTTGCTGAGTCT 351
      ::::: :::::
657 heLysGlyArgThrTyrTrpAsnLeuThrSerLeuAsnValSerGluSer 673
      ::::: :::::
352 .....ACTCTCCAGGTAATGAT..... 369
      ::::: :::::
674 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspSerAlaGlyTh 690
      ::::: :::::
370 .....GAACCTACACT.....GAGCAGA 388
      ::::: :::::
690 rLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThr 707
```

389 TGGCAACTGTGGAGAGA.....AGGGTGAAGAGATCCACCTCTACCTCT 432
 ||| ||||| ||||| ||||| : : : : :
 707 hrPheAsn.ValGluArgAsnAlaArgValAsnPheAspIleLysAlaPr 723
 433 GATTTTCATTCAGCAAAAAAGTTAGCTGATATATGGACCAACAAGTAGGC 482
 ||||: : : : : ||| ||||| : : : : :
 723 oileGlyIleAsnLysTyrSerSerLeuAsnTyrAlaSerPheAsnGlyA 740
 483 ATTGTTCATCAGTGGCTCATCTACGATGGGAGTATTGGACGAGPACA 532
 : : : : : ||||| ||||| : : : : :
 740 snIleSerValSerGly.....GlyGlySerValAspPheThr 752
 533 ATAATGATGAGAAATTCCTACTTATCCAAATGGAAGAATACA.....AGCA 576
 : : : : : : : : : : ||||| : : : : :
 753 LeuLeuAlaSerSerSerAsn.ValGlnThrProGlyValValIleAsnS 769
 577 GTAAGATGTTCCAGCAGGTATTACGTGGTACAAATGTACTAAGAAGTCTCA 626
 ||||| ||||: : : : : ||||| : : : : :
 769 erLysTyrPheAsn.ValSerThrGlySerSerLeuArgPheLysThrSe 785
 627 GGG.....GGCAGCTGTTACACAAAAAGATGCACATCCAAT. 663
 : : : : : ||||| ||||| : : : : :
 785 rGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeuAsnA 802
 664AAAGTAACAGGACTCTATGAAAAGGATGTGAGTTGTT 702
 : : : : : ||||| ||||| : : : : :
 802 laThrGlyLysAsnIleThrLeuGlnValGluGlyThrAspGlyMet 818
 703 CTCCAATCCGCCAGAGCGGAGAGCGTTCATATAATGTTT.....741
 : : : : : ||||| ||||| : : : : :
 819 IleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyGlyAs 835
 742GCACACATGTTGATCTCATAGTGTG 766
 835 nIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsnValT 852
 767 AATCTCTGTACAGA.....CAAAC 786
 : : : : : ||||| : : : : :
 852 hrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPheAspAsn 868
 787 CACAACAAGAAGCTCCAAACAAGCAAAAT.....CAAAAATGCCAA 827
 ||||: ||||| ||||: : : : : : : : : :
 869 HisGlnLysProLeuThrIleLysLysAspValIleIleAsnSerGlyAs 885
 828 TCTCCGAGCACATGGGAAGTG.....ATCC 853
 ||||| : : : : : : : : : : : : : : :
 885 nLeuThrAlaGlyGlyAsnIleValAlaIleAlaGlyAsnLeuThrValG 902
 854 GTGATCTCGAGGACTTTAAGAAAAACCACTCTCATGACA.....891
 : : : : : ||||| ||||| : : : : :
 902 luSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsnValGly 918
 892ACACAGCCACCAATCCACCTTCCTCATTTGTGCGAGATTGG 932
 : : : : : ||||| : : : : : : : : : :
 919 GlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLysGlyGly 935
 933 ACAAGAATTCGTGTTTAGTCTCTTGCAATCTGGAAGCATCGCGACTG 982
 | ||||: : : : : ||||: : : : : ||||: : : : :
 935 yAlaArgPheLysAspIleAspAsnSerLysAsnLeuSerIleThrThrA 952
 983 GTAACCGGCTC.....AATCGCATCAATCAA 1008
 : : : : : ||||| : : : : :
 952 snSerSerSerThrTyrArgThrIleIleSerGlyAsnIleThrAsnLys 968
 1009 GCAGGCCAGCTTTCCTGCTCGACAGACAGTTGAGCTGGGGCTCGGTGG 1058
 ||||: ||||| : : : : : ||||| : : : : :
 969 AsnGlyAspLeuAsnIleThrAsnGlu.....GlySer.....979
 1059 GATGGTGACATTTGACAGTGTGCGCCATGTACAAAGTGAACATCATACAG. 1107
 ||||: ||||: : : : : : : : : : :
 980AspThrGluMetGlnIleGlyGlyAspValSerGlnL 992

1108ATAAACAGTGGCAGTGTACAGGGACACACTCCGCCAAAAGA 1144
 992 ysGluGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrLysGln 1008
 1147TTACTCGACAGCTTCAGGAGGCGCTCATCTGCAGCGGCT 1190
 1009 IleThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAspAlaTh 1025
 1191 TCGATCGGCATTACTGTATTAGGAAGAAA..... 1221
 1025 rAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnA 1042
 1222TATCCAACTGATGA 1236
 1042 sPLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGly 1058
 1237 TCTGAATTTGCTGCTGACGATGGGAAGACACACTATAAGTGGG... 1284
 1059 SerAsp.....LeuThrIleGlyAsnThrAsnSerAlaAspGlyTh 1072
 1285TGCTTTAACGAGGTCAAAACAAGT.....GGTG 1312
 1072 rAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089
 1313 CCATCATCCACAGTCGCTTTTGGGCCCTCTGCAGCTCAAGAAGTAGAG 1362
 1089 IaAspGlyHisLysValThrLeuHis..... 1097
 1363 GAGCTGCCAAATG...ACAGGAGTTTACAGACATATGCTTCAGATCA 1409
 1098SerLysValGluThrSerGlySerAsnAsnAsnThrGluAspSe 1112
 1410 AGTTCAACAACATGGCCTCATTTGATGCTTTTGGGCCCTTTTCATCAGGAA 1459
 1112 rSerAspAsnAsnAla..... 1117
 1460 ATGGAGCTGTCTCTCAGCGCTCCATPCACAGTTGAGAGTAAGGATTAAAC 1509
 1118GlyLeuThrIleAspAlaLysAsnValThr 1127
 1510 CTCAGACACGCCAGTGGATCAATGGCAGCTGATCGTGGACAGCACCT 1559
 1128 ValAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144
 1560 GGGAAAGGACACTTGTCTTATCACCTCGCAACAGCAGCGCTCCCCAAA 1609
 1144 rGlyGlu.....IleThrLysThr..... 1151
 1610 TCCTTCTCTGGGATCCAGTGGACAGAAAGGTGGCTTTGTAGTGGAC 1659
 1152GlyThrThrIleAsn 1156
 1660 AAAACACCAAAATGGCCTACCTCCAAATCCAGGCATTTGATGGTTGG 1709
 1157 AlaThrThrGly.....AsnValGluIleThrAlaGlnThrGln 1169
 1710 CACTTGGAAATACAGTCTGCAAGCAAGCTCAACACCTTGACCCCTGACTG 1759
 1169 ySerIleLeuGlyGlyIleGluSerSerGlySerValThrLeuThrA 1186
 1760 TCAGTCCCGCTGGC.....TCCAAATGTACCTGCCTCCCAATTACA 1800
 1186 IaThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
 1801 GTCACTTCCAAACACAGGACACACAGC..... 1830
 1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLy 1219
 1830 1830
 1219 sGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyGlyT 1236
 1831AAATTCGCCAGCGCCTCTGGTAGTT 1854

us-09-049-696-18.ra1

1236 hrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThrThr 1252
||| ::: |||:||||:
1855 TATGCAGAAATATTCCG.....CAAGGAGCCTCCCACAAATTCACG 1892
:|||| ::: |
1253 GluSerAsnSerLysLysAlaThrThrGlyGluAlaAsnValThrSe 1269
|
1893 GGCAGGTGTCACAGCCCTGATTGAATCAGTGAATGGAAAACAGTTACT 1942
||||: ||| ::: |||:||||:
1269 rAlaThrGlyThr...IleclYclyThrIleSerGlyAsnThrValAsnV 1285
|
1943 TGGAA.....CTACTGGATAAATGGACGAGGTGCTCAT 1974
: ::: |||:||||:
1285 alThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluLeasn 1301
|
1975 GCYTACTAAG.....GATGACGGTGTTCTACTCAAG 2003
|||||: ::: |
1302 AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 1318
|
2004 GTATTTCCACAACCTATTCACACAGNAAGGTAGATACAGTGTAAAAAGTCGGG 2053
:||||: |||:||||: ::: |||:||||:
1318 rGluAlaSerSerHisIleThrSerAlaLysGlyGlnValAsnLeuSerA 1335
|
2054 CT.....CTGGGAGGAGTTAACCCAGCCAGACGGAGAGTG 2088
|| |||:||||: |||:||||:
1335 laGlnAspGlySerValAlaGlySerIleAsnAlaAla.....AsnVal 1349
|
2089 ATACCCCGACGAGGTGGACACCTGTACATACCTGGCTGGATTACAGAATGA 2138
: ::: |||:||||:
1350 ThrLeuAsnThrThrGlyThrLeuThrThr.....ValLysGlySe 1363
|
2139 TGAATCAATCAATGGAAATCCACACAGACCTGNAATTAATFAGGATGATGTC 2188
:||||: ::: |||:||||:
1363 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 1380
|
2189 AA.....CACAGCAAGTGTGTTTCACAGCA 2214
: ||| |||:||||:
1380 luLeuAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn 1396
|
2215 ACATCTCTGGGAGGCTCATTTGTGGCTTCTGATGTCGCCAAATGCTCCCAT 2264
:||||: |||:||||: ::: |||:||||:
1397 AlaAsnGlySerGlySerValIleAlaThrThrSerSerArgValAsnIl 1413
|
2265 ACCT...GATCTCTTCCACTCGGCCAAATC.....ACCG 2296
|| |||:||||: ::: |
1413 eThrGlyAspLeulleThrIleAsnGlyLeuAsnIleIleSerLysAsnG 1430
|
2297 ACCTGAAGCGGGAATTCACGGGGGCGAGTCTCATTAATCTGACTCTTGACA 2346
|
1430 lYlleAsnThrValLeuLeuLysGlyValLysIleAspValLysTrIle 1446
|
2347 GCTCTCTGG...GATGATTATGACCATTGGAACAGCTCACAAATATATCAT 2393
|||||: ::: |||:||||:
1447 GluProGlyIleAlaSerValAspGluValIleGluAlaLys..... 1460
|
2394 TCGAATAAGTACAAGTATTTCTTGATCTCAGACAGAAGTTCATGAATCTC 2443
|||||: |||:||||: ::: |||:||||:
1461 .ArgileuLeuGlyLysValLysAspLeuSerAspGluArgGluAlaL 1477
|
2444 TT 2445
||

```
alignment_scores:
  Quality: 138.50      Length: 936
  Ratio: 0.318        Gaps: 49
  Percent Similarity: 46.581  Percent Identity: 20.406
```

```
alignment_block:
US-09-049-696-18 x US-08-302-832-2 ..
```

Align seq 1/1 to: MS-08-302-832-2 from: 1 to: 1536

217' TTTTGAAGCTACA.....GGAAAGCGATTTTATTTCACAAATGT 254
 |||||:::||||
 631' PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnIle 643
 |||||:::||||
 255' TGCCATTITGATCTCTGAAACATGGAGACAAAGGCTGACTATCTGAGAC 304
 |||||:::||||
 643' eserMetValLeuPro.....LysAsnGluSerGlyTyrrAspLysP 657
 |||||:::||||
 305' CAAAACTTCGAGACCTACAAAAATGCTGATGTTCTG...GTTGCTGAGTCT 351
 |||||:::|||||
 657' heLysGlyArgThrTyrrAsnLeuThrSerLeuAsnValSerGluSer 673
 |||||:::|||||
 352'ACTCCTCCAGGTGAATGAT..... 369
 |||||:::||||
 674' GlyGluPheAsnLeuThrIleAspSerArgGlySerAspSerAlaGlyTh 690
 |||||:::||||
 370'GAACCTACACT.....GAGCAGA 388
 |||||:::||||
 690' rLeuThrGlnProTyrrAsnLeuAsnGlyIleSerPheAsnLysAspThr 707
 |||||:::||||
 389' TGGGCAACTGTGCAGAGA.....AGGCTGAAGGATCCCACTCACTCT 432
 |||||:::||||
 707' hePheAsnValGluArgAsnAlaArgValAsnPheAspIleLysAlaPr 723
 |||||:::||||

seq name: /can2 6/ptodata/2/iaa/5A COMB.pep:US-08-302-832-2

```

seq_documentation_block:
: sequence 2, Application US/08302832
: Patent No. 5603938
: GENERAL INFORMATION:
:
: APPLICANT: Barenkamp, Stephen J
:
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
:
: DATE OF INVENTION: No. 5603938-Typeable Haemophilus
:
: NUMBER OF SEQUENCES: 8
:

```



```

1253 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 1269
1293 GGCCAGTGTTCACACCCCTGATGTAATGAGTAAATGGAACACATTTACCT 1942
1269 rAlaThrGlyThr...IleGlyGlyThrIleSerGlyAsnThrValAsnV 1285
1943 TGGAA.....CTACTGGTAATGAGCAGAGGTGCTGAT 1974
1285 aiThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluAlaAsn 1301
1975 GCTACTAAG.....GATCAGCGTGTCTACTCAAG 2003
1302 AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 1318
2004 GTATTTCACACTTATGACAGCAATGTTAGTACAGTGTAAAGTCGGG 2053
1318 rGluAlaSerSerHisIleThrSerAlaLysGlyGlnValAsnLeuSerA 1335
2054 CT.....CTGGGAGAGTTAACGCGCAGCCAGCAGGAGGTG 2088
1335 laGlnAspGlySerValAlaGlySerIleAsnAlaAla.....AsnVal 1349
2089 ATACCCAGCAGAGTGGAGCACTGTACATACCTGGCTGGATTGAGAATGA 2138
1350 ThrLeuAsnThrThrGlyThrLeuThr.....ValLysGlySe 1363
2139 TGAATACATGGAATCCACCAAGCTGAAATTAATAAGATGATGTC 2188
1363 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 1380
2189 AA.....CACAGCAAGTGTGTTTCAGCAGA 2214
1380 luLeuAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn 1396
2215 ACATCTCGGAGGCTCATTTGTGGCTTCTGATGTCCCAATGCTCCCAT 2264
1397 AlaAsnGlySerGlySerValIleAlaThrThrSerSerArgValAsnI 1413
2265 ACCT...GATCTCTCCACCTGCGCCAAATC.....ACCG 2296
1413 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleIleSerLysAsnG 1430
2297 ACCTGAAGGGGAAATCACGGGGCAGCTCTCATTATCTCAGCTTGAC 2346
1430 lyIleAsnThrValLeuLeuLysGlyValLysIleAspValLysIle 1446
2347 GCTCCTGGG...GATGATTATGACCATGGAAACAGCTCACAAATATATCAT 2393
1447 GlnProGlyIleAlaSerValAspGluValIleGluAlaLys..... 1460
2394 TCGAATAGTACAGTATTTCTGATCTCAGAGCAAGTTCAATGAATCTC 2443
1461 ArgIleLeuGluLysValLysAspLeuSerAspGluArgGluAlaL 1477
2444 TT 2445
1477 eu 1477

```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:us-08-530-198-2

```

seq_documentation_block:
; Sequence 2, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCE: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

```

```

STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,198
FILING DATE: 13-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: JWB-1186
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-530-198-2

```

```

alignment_scores:
Quality: 138.50      Length: 936
Ratio: 0.318        Gaps: 49
Percent Similarity: 46.581    Percent Identity: 20.406
alignment_block:
US-09-049-696-18 x US-08-530-198-2

```

```

Align seg 1/1 to: US-08-530-198-2 from: 1 to: 1536
217 TTTGAAGCTACA.....GGAAGCGATTTTATTTCAAAAATGT 254
631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnI 643
255 TGCCATTTTGTATCTCAACATGGAAGCAAGGCTGACTATGTGAGAC 304
643 eSerMetValLeuPro.....LysAsnGluSerGlyTyrAspLysP 657
305 CAACACTTGAGACCTACAAAATGCTGATGTTCTG....GTTGCTGAGTCT 351
657 heLysGlyArgThrTyrTrpAsnLeuThrSerLeuAsnValSerGluSer 673
352 .....ACTCCTCCAGGTAATGAT..... 369
674 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspSerAlaGlyTh 690
370 .....GAACCCCTACACT.....GAGCAGA 388
690 rLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThr 707
389 TGGGCAACTGTGGAGAGA.....AGGGTGAAGGATCCACCTCCTCCT 432
707 hrPheAsn.ValGluArgAsnAlaArgValAsnPheAspIleLysAlaPr 723
433 GATTTCATTGCAGGAAAAAATAGCTGAATATGGACCACAAAGGTAGGCG 482
723 oileGlyIleAsnLysTyrSerSerLeuAsnTyrAlaSerPheAsnGlyA 740
483 ATTTGTCCATGAGTGGGCTCATCTACGATGGGAGTATTTGACGAGTACA 532
740 snIleSerValSerGly.....GlyGlySerValAspPheThr 752

```


[illegible]

1042 splLeuAsnIleSerGlyPheAsnIysAlaGluIleThrAlaLysAspGly 1058
1237 TCTGAATTGTCTCTCTCAGCGATGGGAACAACTATAAGTGG... 1284
1059 SerAsp.....LeuThrIleGlyAsnThrAsnSerAlaAspGlyTh 1072
1285TGCCTTTAAGAGGTCAACAAGT.....GGTG 1312
1072 rAsnAlaIysLysValThrPheAsnGlnValLysAspSerLysLeu 1089
1313 CCATCATCCACACAGCTCGCTTTTGGGGCCCTCTGCAGCTCAAGAACATAGAG 1362
1089 laAspGlyHisLysValThrLeuHis..... 1097
1363 GAGCTGTCCTCAAAATG..ACAGGAGGTTCACAGACATATGCTTCAGATCA 1409
1098SerLysValGluThrSerGlySerAsnAsnAsnThrGluAspSe 1112
1410 AGTTCAAGAACATGGCCCTCATGTGCTTTTGGGGCCCTTCATCAGAA 1459
1112 rSerAspAsnAla..... 1117
1460 ATGGAGCTGCTCTCTCAGCGCTCCATCCAGCTTGAGACTAAGGATTAACC 1509
1118GlyLeuThrIleAspAlaLysAsnValThr 1127
1510 CTCCAGACAGCAGCTGGATGAATGGACAGCTGATCTGTGCACAGCACCGT 1559
1128 ValAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144
1560 GGGAAAGGACACTTGTGTTCTTATCATCGTGACACAGCAGCCCTCCCCAAA 1609
1144 rGlyGlu.....IleThrLysThr..... 1151
1610 TCCTTCTCTGGATCCAGTGGACAGACAGAGTGCTGTTAGTGGAC 1659
1152GlyThrIleAsn 1156
1660 AAAAAACGCAAAATGGCTACTCTCCAATCCAGGCATGCTAAGTTGG 1709
1157 AlaThrThrGly.....AsnValGluIleThrAlaGlnThrGl 1169
1710 CACTTGGAAATACAGTCTGGAAGCAGCTCACAACTTGACCTGACTG 1759
1169 ySerIleLeuGlyGlyIleGluSerSerGlySerValThrLeuThrA 1186
1760 TCAGTCCCGTGC.....TCCAATGTCACCTGCCTCCAAATTACA 1800
1185 laThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
1801 GTGACTTCCAAAACGAACAGCACCAAGC..... 1830
1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLy 1219
1830 1830
1219 sgLysThrGluSerValThrThrSerSerGlnSerGlyAspIleGlycylt 1236
1831AAATGCCCGACCCCTCTGTAGTGT 1854
1236 hrIleSerGlycylThrValGluValLysAlaThrGluSerLeuThrThr 1252
1855 TATGCAAAATATTCG.....CAAGGAGCCTCCCAATCTTCAG 1892
1253 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 1269
1893 GGCAGTGTCAACCCCTGATTGTAATCAGTCAATGGAAAAACAGTTACCT 1942
1269 rAlaThrGlyThr.....IleGlyGlyThrIleSerGlyAsnThrValAsnV 1285
1943 TGGAA.....CTACTGGATATGAGCAGCGTCTGAT 1974

753 GTAAGATGTTTCAGCAGGATTACTCGTACAAATAGTAGAAGAGTGTCA 626
757 erLysTy-PheAsn.ValSerThrGlySerSerLeuArgPheLysThrSe 785
627 GGGG.....GGCAGCTGTTACACCAAAAGATGCACATTCAT. 663
785 rGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeuAsn 802
664AAAGTAACAGGACTCTATGAAAAGATGTGAGCTTTGTT 702
802 laThrGlyLysnIleThrLeuGlnValGluGlyThrAspGlyMet 818
703 CTCCAATCCGCCAGACAGGAGGCTTCTATAATGTTT..... 741
819 ileGlyLysGlyIleValAlaLysAsnIleThrPheGluGlyGlyAs 835
742GCACACATGTTGATTCATCTATAGTTG 766
835 nileThrPheGlySerArgLysAlaValThrGluIleGluGlyAsnVal 852
767 AATCTGTGTACAGAA.....CAAAAC 786
852 hrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPheAspAsn 868
787 CACAACAAGAGCTCCAAACAGCAAAAT.....CAAAATGCCAA 827
869 HisGlnLysProLeuThrIleLysLysAspValIleAsnSerGlyAs 885
828 TCTCGAAGCACATGGGAAGTG.....ATCC 853
885 nLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuThrValG 902
854 GTGATTCGAGGACTTTAAGAAAACCACTCCTATGACA..... 891
902 LuSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsnValGly 918
892ACAGCCACCACCAATCCCACTTCATTCGTGCTGAGATGG 932
919 GlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLysGlyG 935
933 ACAAGAATTTGCTTTAGTCTCTTGACAAATCTGGAGCATGCGCACTG 982
935 yAlaArgPheLysAspIleAspAsnSerLysAsnLeuSerIleThrThr 952
983 GTAACCGCCTC.....AATCGACTGAATCAA 1008
952 snSerSerThrTyrArgThrIleIleSerGlyAsnIleThrAsnLys 968
1009 GCAGGCGAGCTTTCTGCTGCACACAGCTTCAGCTGGGTCTGGTGG 1058
969 AsnGlyAspLeuAsnIleThrAsnGlu.....GlySer..... 979
1059 GATGTGACATTTGACGTGCTGCCATGTACAAAGTCAACTCATACAG. 1107
980AspThrGluMetGlnIleGlyGlyAspValSerGlnL 992
1108ATAACAGTGGCAGTGCAGGGACACACATCGCCAAAAGA 1146
992 ysGluGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrLysGln 1008
1147TTACCTGCAGCAGCTTCAGGAGGACGTCATCCACGCGGCT 1190
1009 IleThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAlaThr 1025
1191 TCGATCGCATTTACTGTGATTAGAGAGAA..... 1221
1025 rAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnA 1042
1222TATCCAACGTGATGGA 1236

104 sLeuAsn1IleSerGlyPheAsn1LysAlaGlu1IleThrAlaLysAspGly 1058
1237 TCTGAATTTGCTGCTGACGGATCGGGAACACAACTATAAGTGG... 1284
1238 TCTGAATTTGCTGCTGACGGATCGGGAACACAACTATAAGTGG... 1284
1239 TCTGAATTTGCTGCTGACGGATCGGGAACACAACTATAAGTGG... 1284
1059 SerAsp...LeuThrIleGlyAsnThrAsnSerAlaAspGly 1072
1285 TCTTTTACGAGGCTCAACAAGT...GGTG 1312
1286 TCTTTTACGAGGCTCAACAAGT...GGTG 1312
1072 rAsnAlaLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089
1313 CCATCATCCACACAGCTCGCTTTGGGGCCCTCTGCAGCTCAAGAAGTACAG 1362
1314 CCATCATCCACACAGCTCGCTTTGGGGCCCTCTGCAGCTCAAGAAGTACAG 1362
1089 laAspGlyHisLysValThrLeuHis... 1097
1363 GAGCTGTCCTCAAAATG...ACAGGAGGTTTACAGACATATCTTCAGATCA 1409
1364 GAGCTGTCCTCAAAATG...ACAGGAGGTTTACAGACATATCTTCAGATCA 1409
1098 SerLysValGluThrSerGlySerAsnAsnThrGluAspSe 1112
1410 AGTTTCAGAACAAATGGCCCTCATGTGCTTTTGGGGCCCTTTCATCAGGAA 1459
1112 rSerAspAsnAla... 1117
1460 ATGGAGCTGCTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAACC 1509
1118 GlyLeuThrIleAspAlaLysAsnValThr 1127
1510 CTCGAGAACAGCCAGTGGATGAATGGCACAGTGTGTCGGACAGCACCGT 1559
1128 ValAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144
1560 GGGAAAGGACACTTGTGTTCTTATCAGCTGCACAACGACGAGCTCCCCAAA 1609
1147 rGlyGlu...IleThrLysThr... 1151
1610 TCCTCTCTCTGGATCCAGTGGACAGAGAAGTGGCTTGTAGTGGAC 1659
1152 GlyThrThrIleAsn 1156
1660 AAAACACACAAATGGCTACTCTCAATCCAGGCATGCTAAGTGTGG 1709
1157 AlaThrThrGly...AsnValGluIleThrAlaGlnThrGly 1169
1710 CACTTGGAAATACAGCTCTGCAAGCAAGCTCACAACTTGACCTGACTG 1759
1169 ySerIleLeuGlyGlyIleGluSerSerGlySerValThrLeuThrA 1186
1760 TCACGTCCTCGCTGG...TCCAAATGCTACCTGCCTCCCAATFACA 1800
1186 IaThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
1803 CTCACCTTCCAAACAGCAAGACACACAGC... 1830
1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLys 1219
1830 sGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGly 1236
1831 AAATTCCTCCAGCCCTCTGGTAGTT 1854
1236 hrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThrThr 1252
1855 TATGCAAAATATTGCG...CAAGGAGCCCTCCCAATTCCTCAG 1892
1253 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 1269
1893 GGCAGTGTCAAGCCCTCATTTGAATCAGTGAATGGAAACACAGTTACCT 1942
1269 rAlaThrGlyThr...IleGlyGlyThrIleSerGlyAsnThrValAsnV 1285
1943 TCGAA...CTACTGGATTAATGGACAGGCTGCTGAT 1974
1285 aThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyValaGluIleAsn 1301

```
1975 GCTACTAAG.....CATGCGGTGCTCTACTCAAG 2003
|||||...
1302 AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 1318
|||||...
2004 CTATTTCACAACTTATGACACAGAAATGGTATACAGTGTAAAGTGGGG 2053
|||||...
1318 rGluAlaSerSerHisIleThrSerAlaLysGlyClnValAsnLeuSerA 1335
|||||...
2054 CT.....CTGGAGGAGTTAAGCAGCAGCAGCGAGAGTG 2088
|||||...
1335 laGlnAspGlySerValAlaGlySerIleAsnAlaAla.....AsnVal 1349
|||||...
2089 ATACCCAGCAGCAGGAGTGGAGCTGTATACATCGCTGGATGAGATGCA 2138
|||||...
1350 ThrLeuAsnThrThrGlyThrLeuThrThr.....ValLysGlySe 1363
|||||...
2139 TGAATACAAATGGAAATCCACCAAGACCTGAAATTAATAAGATGATGTC 2188
|||||...
1363 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 1380
|||||...
2189 AA.....CACAAGCAAGTGTGTTTCAGACAGA 2214
|||||...
1380 luLeuAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn 1396
|||||...
2215 ACATCCTCGGAGGCTCATTTGTGGCTTCTGATGTCCTCAAAATGTCCTCC 2264
|||||...
1397 AlaAsnGlySerGlySerValIleAlaThrThrSerSerArgValAsnIl 1413
|||||...
2265 ACCT...GATCTCTCCACCTGGCGCAATC.....ACCG 2296
|||||...
1413 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleSerLysAsnG 1430
|||||...
2297 ACCTGAAGCGGAAATTCACGGGCGAGCTCTCATTAATCTGACTGTGACA 2346
|||||...
1430 lyIleAsnThrValLeuLeuLysGlyValLysIleAspValLysTyrIle 1446
|||||...
2347 GCTCTCTGG...GATGATTATGACCATGGAACAGCTCACAGTATATCAT 2393
|||||...
1447 GlnProGlyIleAlaSerValAspGluValIleGluAlaLys..... 1460
|||||...
2394 TCGAATAAGTACAACTATCTGATCTCAGACACAAAGTTCATGATGATCTC 2443
|||||...
1461 ArgIleLeuGluLysValLysAspLeuSerAspGluArgGluAlaL 1477
|||||...
2444 TT 2445
||
1477 eu 1477
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-728-470-2

seq_documentation_block:

; Sequence 2, Application US/08728470
; Patent No. 5928651

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins

; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESS: Shoemaker and Mattare, Ltd.

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202-0286

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; US-08-728-470-2

alignment_scores:

Quality: 138.50 Length: 936
Ratio: 0.318 Gaps: 49
Percent Similarity: 46.581 Percent Identity: 20.406

alignment_block:

US-09-049-696-18 x US-08-728-470-2 ..

Align seg 1/1 to: US-08-728-470-2 from: 1 to: 1536

```
217 TTTGAAGCTACA.....GGAAGCGATTTTATTTCAAAAATGT 254
|||||...
631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnIl 643
|||||...
255 TCCCATTTTGATTCCTGAACATGGAACAAAGCTGACTATGTGAGAC 304
|||||...
643 eSerMetValLeuPro.....LysAsnGluSerGlyTyrAspLysp 657
|||||...
305 CAAAACCTTGAGACCTACAAAATGCTGATGTTCTG...GTTGCTGAGTCT 351
|||||...
657 heLysGlyArgThrTyrTrpAsnLeuThrSerLeuAsnValSerGluSer 673
|||||...
352 .....ACTCTCCAGGTAATGAT..... 369
|||||...
674 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspAlaGlyTh 690
|||||...
370 .....CAACCCCTACACT.....GAGCAGA 388
|||||...
690 rLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThr 707
|||||...
389 TGGGCAACTGTGGAGAGA.....AGGTGAAAGATCCACTCCTCCT 432
|||||...
707 hrPheAsn.ValGluArgAsnAlaArgValAsnPheAspIleLysAlaPr 723
|||||...
433 GATTTCATTGCGAGGAAAAAGTTAGCTCAATATGGACCACAGGTAGGCG 482
|||||...
723 oileGlyIleAsnLysTyrSerSerLeuAsnTyrAlaSerPheAsnGlyA 740
|||||...
483 ATTGTCCTCATGAGTGGGCTCATCTACGATGGGAGTATTTCACGAGTACA 532
|||||...
740 snIleSerValSerGly.....GlyGlySerValAspPheThr 752
|||||...
533 ATAATGATGAGAAATTCATCTACATTCCAATGGAAGATACA.....ACGA 576
|||||...
|||||...
|||||...
```

753 IeuLeuAlaSerSerSerAsn...ValGlnThrProGlyValValIleAsnS 769

577 GTAAGATGTTCACGAGGTATTACTGGTACAATCTAGTAAGAAGAGTGCA 626
||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :

769 erLysTyRPhEasN.ValSerThrGlySerSerLeuArgPheLysThrSe 785

627 GGGG.....GGCAGCTGTTACACCACAAAGATGCATTCAT.. 663
:||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

785 rGLySerThryLysThrGlyPheSerIleGLuLyAspLeuThrLeuAsnA 802

664AAAGTAAACAGCATCTATGAAGAAGATGTGACTTTGTT 702

802 laThrGlyLysAsnIleThrLeuGlnValGluGlyThrAspGlyMet 818

703 CTCCAATCCCGCACGAGAGGCTCTATAATGTTT..... 741

819 IleGlyLysGlyLeValalalylsAsnIleThrPheGluGlyGlyAs 835

742GCACACATGTTGATTCTATCATATTG 766

835 ntleThrPheGlySerArGLysAlaValThrGlulleGLuGLyAsnValt 852

767 AATCTGTGTACAAA.....CAAAC 786

852 hrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPheAspAsn 868

787 CACAACAAGAGCTCCAACAAGCAAAAT.....CAAAAATGCCAA 827

869 HisGlnLysProLeuThrIleLysLysAspValIleIleAsnSerGlyAs 885

828 TCTCGAGACACATGGGAAGTG.....ATCC 853
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

885 nLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuThrValG 902

854 GTGATTCTGAGGACTTAAGAAACCCTCCTATGACA..... 891

902 luserAsnAlaAsnPhelysAlaIleThrAsnPheThrPheAsnValgly 918

892ACACGCCAACCAATCCCACCTTCATTGTGTCGACATGG 932
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

919 GlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLysGlyGI 935

933 ACAAGAATTTGCTTTAGTCTCTTGACAATCTGGAAGCATGCGCACT 982
| ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

935 yAlaargPheLysAspIleAspAsnSerLysAsnLeuSerIleThrThrA 952

983 GTAACCGCCTC.....AATCGACTCAATCAA 1008
:|||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :||||

952 snSerSerSerThrTyRArgThrIleIleSerGlyAsnIleThrAsnLys 968

1009 GCAGGCCAGCTTTTCCTGTCGACACAGATTGAGCTGGGTCTGGTGG 1058
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

969 AsnGlyAspLeuAsnIleThrAsnGlu.....GlySer..... 979

1059 GATGTGACATTTGACAGTGTGCCCCATGTACAAGTGAACATCATACAG. 1107
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

980AspThrGluMetGlnIleGlyGlyAspValSerGlnL 992

1108ATAACAGCTGGCAGTGACAGGGACACATCGCCCAAGA 1146
:|||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :||||

992 ysGluGlyAsnLeuThrIleSerSerAspylIleAsnIleThrLysGln 1008

1147TTACCTGACAGCTTCAGAGGGACGTCCTATCTGCACGGGCT 1190
:|||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :||||

1009 IleThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAspAlaTh 1025

1191 TCGATTCGGCATTTACTGTGATTAGGAAGAA..... 1221
:|||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :||||

1025 rAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnA 1042

1222TATCCAACGTGATGGA 1236
:|||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :|||| :||||

1042 spLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGly 1058

1237 TCTGAATTTGTCTGCTCAGCGATTGGGAACACAACACTATAAAGTGG... 1284
1238 : : : : : LeuThrIleGlyAsnThrAsnSerAlaAspGlyThr 1072
1059 SerAsp.....LeuThrIleGlyAsnThrAsnSerAlaAspGlyThr 1072

1285 : : : : : TGCTTTAACGAGGTCAACAAGAAT.....GGTG 1312
1286 :
1072 rAsnAlaLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089
1313 CCATCATCCACACAGTCGCCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAG 1362
1314 :
1089 laAspGlyHisLysValThrLeuHis..... 1097

1363 GAGCTGTCCTCAAATG...ACAGGAGGTTTACAGACATATGCTTCAGATCA 1409
1364 :
1098SerLysValIcLuThrSerGlySerAsnAsnThrGluAspSe 1112
1410 AGTTCAGACAATGCGCTCATGTGCTTTTGGGGCCCTTTCATCAGAA 1459
1411 :
1112 rSerAspAsnAsnAla..... 1117

1460 ATGAGCTGCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGGATTAAACC 1509
1461 :
1118GlyLeuThrIleAspAlaLysAsnValThr 1127
1510 CTCCAAGACGCCATGGATGAATGGCACAGTAGTCGTGGACAGCACCGT 1559
1511 :
1128 ValAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144
1560 GGGAAAGGACACTTGTTCCTATCACCTGCACACGACGCGCTCCCACAA 1609
1561 :
1144 rGlyGlu.....IleThrLysThr..... 1151
1610 TCCTCTCTGGCATCCAGTGCAGAGAAGGTGGCTTTGTAGTGCAG 1659
1611 :
1152GlyThrThrIleAsn 1156
1660 AAAAACACCAAAATGCGCTACTCTCCAAATCCAGGCATTCGTAAAGTTGG 1709
1661 :
1157 AlaThrThrGly.....AsnValGluIleThrAlaGlnThrGl 1169
1710 CACTTGGAAATACACTCTGCAGCAGCAGCTCAAAACCTTGACCTGACTG 1759
1711 :
1169 ySerIleLeuGlyIleGluSerSerSerGlySerValThrLeuThrA 1186
1760 TCACGTCGCGTCG.....TCCAAATGCTACCTGCCTCCCAATTACA 1800
1761 :
1186 lathrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
1801 GTGACTTCCAAAACGAACAGACACACAGC..... 1830
1802 :
1203 valThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLy 1219
1830 1830
1219 sGlyThrGluSerValThrThrSerSerGlnSerGlyAspilegLy 1236
1831AAATTCGCCGCCCTCTCGTAGTT 1854
1236 hrIleSerGlyThrValGluValLysAlaThrGluSerLeuThrThr 1252
1855 TATGCAATATTTCGC.....CAAGGACGCTCCCAATCTTCAG 1892
1856 :
1253 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 1269
1893 GGCCAGTGTCAAGCCCTGATTGAATCAGTGAATGGAAAAACAGTTACCT 1942
1269 rAlaThrGlyThr....IleGlyGlyThrIleSerGlyAsnThrValAsnV 1285
1943 TGGAA.....CTACTCGGATAATGGACAGCGGTGCTGAT 1974
1285 alThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyValGluIleAsn 1301


```

769 erLysThrPheAsn.ValSerThrGlySerSerLeuArgPheLysThrSe 785
627 GGGA.....GGCAGCTGTTACACCAAAAGATGCACATCAAT. 663
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
785 rGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeuAsnA 802
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
664 .....AAAGTAACAGGACTCTATCAAAAGGATCTGAGTTGTT 702
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
802 laThrGlyGlyAsnIleThrLeuGlnValGluGlyThrAspGlyMet 818
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
703 CTCCAATCCGCCAGAGAGAGGCTTCTATATGTT..... 741
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
819 IleGlyGlyGlyIleValAlaLysLysAsnIleThrPheGluGlyGlyAs 835
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
742 .....GCACAACATGTTGATCTATAGTTG 766
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
835 nileThrPheGlySerArgLysAlaValThrGluIleGluGlyAsnVal 852
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
767 AATCTGTACAGAA.....CAAAAC 786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
852 hrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPheAspAsn 868
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
787 CACAACAAGAAGCTCAACAAGCAAAAT.....CAAAAATGCAA 827
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
869 HisGlnLysProLeuThrIleLysLysAspValIleAsnSerGlyAs 885
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
828 TCTCCGAAGCACATGGGAAGTG.....ATCC 853
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
885 nLeuThrAlaGlyGlyAsnIleValAlaGlyAsnLeuThrValG 902
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
854 GTGATTCTGAGGACTTAAGAAACACCTCTATGACA..... 891
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
902 luSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsnValGly 918
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
892 .....ACAGCACCAAAATCCACCTTCTCATTTGTCGAGATGG 932
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
919 GlyLeuPheAspAsnLysGlyAsnSerAsnIleAlaLysGlyGI 935
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
933 ACAAAAGATGTGTTAGTCTCTGACAAATCTGGAAGCATGGCAGTG 982
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
935 yAlaArgPheLysAspIleAspAsnSerLysAsnLeuSerIleThrA 952
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
983 GTAACCGCCTC.....AATCGACTGAATCAA 1008
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
952 snSerSerThrThrArgThrIleIleSerGlyAsnIleThrAsnLys 968
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1009 GCAGCCAGCTTTCTCTGCTGCAGACAGTTGAGCTGGGTCTGGTTG 1058
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
969 AsnGlyAspLeuAsnIleThrAsnGlu.....GlySer..... 979
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1059 GATGGTGACATTTGACAGTGCTCCCATGTACAAAGTGAATCATACAG. 1107
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
980 .....AspThrGluMetGlnIleGlyGlyAspValSerGlnL 992
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1108 .....ATAACAGTGGCAGTGACAGGCACACACTCGCCAAAAGA 1146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
992 ysGluGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrLysGln 1008
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1147 .....TTACCTGCAGCAGCTTCAGAGGACGCTCCATCTGCAGCGGCT 1190
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1009 IleThrIleLysAlaGlyValAspGlyGluAsnSerAspAlaLath 1025
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1191 TCATCGGCAATTTACTGTGATTAAGGAAGAA..... 1221
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1025 rAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnA 1042
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1222 .....TATCCAACCTGATGGA 1236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1042 spLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGly 1058
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1237 TCTGAAATGTGCTGCTGCGGATGGGAAGACACACTATAAGTGGG.. 1284
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1059 SerAsp.....LeuThrIleGlyAsnThrAsnSerAlaAspGlyTh 1072

```

```

1285 .....TGCTTTAACGAGGCTCAACAAGT.....GGTG 1312
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1072 rAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1312 CCATCATCCACACAGCTGCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAG 1362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1089 laaspGlyHisLysValThrLeuHis..... 1097
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1363 GAGCTGTCCAAAATG...ACAGGAGGTTTACAGACATATGCTTCAGATCA 1409
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1098 .....SerLysValGluThrSerGlySerAsnAsnThrGluAspSe 1112
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1410 AGTTCAGAACATGGCTCATGTGCTTTTGGGGCCCTTTTCATCAGGAA 1459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1112 rSerAspAsnAla..... 1117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1460 ATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGACAGTAAGGGATTAAACC 1509
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1119 .....GlyLeuThrIleAspAlaLysAsnValThr 1127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1510 CTCGAGAACAGCCAGTGGATGAATGGCACAGTATCGTGACAGACACCGT 1559
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1128 ValAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1560 GGGAAGGACACTTTGTTTCTTATCCTCGACAAACGACGCTCCCAAA 1609
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1144 rGlyGlu.....IleThrThrLysThr..... 1151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1610 TCCTTCTCTGGATCCAGTGGACAGAGAGGCTGGCTTTGTAGTGGAC 1659
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1152 .....GlyThrThrIleAsn 1156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1660 AAAAACACCAAAATGGCTACTCTCCAAATCCCGGCATTTGCTAAGTTGG 1709
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1157 AlaThrThrGly.....AsnValGluIleThrAlaGlnThrGI 1169
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1710 CACTTGGAAATACAGTCTGCAACCAAGCTCACAACCTTGACCTGACTG 1759
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1169 ySerIleLeuGlyGlyIleGluSerSerGlySerValThrLeuThrA 1186
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1760 TCAGTCCCGCTGCG.....TCCAATGCTACCTGCTCCCAATTACA 1800
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1186 laThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1801 GTGACTTCCAAAACGAACAGGACACACG..... 1830
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLy 1219
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1830 ..... 1830
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1219 sGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyLT 1236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1831 .....AAATTCCCGCCAGCTCTCGTAGTT 1854
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1236 hrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThrThr 1252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1855 TATGCAAAATATTCG.....CAAGGAGCTCTCCCAATTCCTCAG 1892
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1253 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 1269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1893 GGCAGTGTACACCCCTGATTGAATCAGTGAATGGAAACACAGTTACCT 1942
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1269 rAlaThrGlyThr...IleGlyGlyThrIleSerGlyAsnThrValAsnV 1285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1943 TGGAA.....CTACTGGATAATGGAGCAGGTGCTCAT 1974
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1285 alThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 1301
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1975 GCTACTAAG.....GATGACGCTGTCTACTCAAG 2003
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1302 AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 1318

```

```
2004 GTATTTCACAACTTATGACAGCAATGTGTAGATACAGTGTATAAAGTGGCGG 2053
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1318 rGluAlaSerSerHisIleThrSerAlaLysGlyGlnValAsnLeuSerA 1335
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2054 CT.....CTGGGAGGAGTTAAGCAGCAGCAGCAGCAGAGTG 2088
      || : : : : : : : : : : : : : : : : : : : : : : : :
1335 laGlnAspGlySerValAlaGlySerIleAsnAlaAla.....AsnVal 1349
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2089 ATACCCAGCAGCAGGAGCAGTGTATACATCTGCTGGTGAATGAGAATGA 2138
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1350 ThrLeuAsnThrThrGlyThrLeuThr.....ValLysGlySe 1363
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2139 TGAATACAAATGGAATCACCAGACCTGAAATTAATAAGATGATGTC 2188
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1363 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 1380
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2189 AA.....CACAAGCAGTGTGTTTCAGACAGA 2214
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1380 luLeuAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn 1396
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2215 ACATCTCGGAGGCTCATTTGTGCTTCTGTATGTCCTCCAAATGCTCCCAT 2264
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1397 AlaAsnGlySerGlySerValIleAlaThrThrSerSerArgValAsnI 1413
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2265 ACCT...GATCTCTCCACCTGGCCAAATC.....ACCG 2296
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1413 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleIleSerLysAsnG 1430
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2297 ACCTGAAGCGCAAAATCAGGGCGCAGTCTCATTAATCTGACTTGGACA 2346
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1430 lyIleAsnThrValLeuLeuLysGlyValLysIleAspValIleThrIle 1446
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2347 GCTCTGGG...GATCATATATGACCATGACAGCTCACAGTATATATCAT 2393
      ||||| : : : : : : : : : : : : : : : : : : : : : : : :
1447 GlnProGlyIleAlaSerValAspGluValIleGluAlaLys..... 1460
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2394 TGAATAAGTACAGTATCTTGATCTCTAGACAGCAAGTCAATGATCTC 2443
      ||||| : : : : : : : : : : : : : : : : : : : : : : : :
1461 rArgIleLeuGluLysValLysAspLeuSerAspGluGluArgGluAlaL 1477
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2444 TT 2445
      ||
1477 eu 1477
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.US-08-719-641-2

```
seq_documentation_block:
; Sequence 2, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
```

```

; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; (NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; US-08-719-641-2

alignment_scores:
      Quality: 138.50      Length: 936
      Ratio: 0.318        Gaps: 49
      Percent Similarity: 46.581      Percent Identity: 20.406

alignment_block:
US-09-049-696-18 x US-08-719-641-2 ..
Align seg 1/1 to: US-08-719-641-2 from: 1 to: 1536
217 TTTGAAGCTACA.....CGAAGCGATTTTATTTCAAAAATGT 254
      ||||| : : : : : : : : : : : : : : : : : : : : : :
631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnI 643
255 TGCATTTTGTATCTCTGAACATGGAACACAAAGCTGACTATGTGACAC 304
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
643 eSerMetValLeuPro.....LysAsnGluSerGlyTyrAspLysP 657
305 CAAAACTTGAGACCTACAAAAATGCTGATGTTCTG...GTTGCTGAGTCT 351
      ||| : : : : : : : : : : : : : : : : : : : : : :
657 heLysGlyArgThrTyrTrpAsnLeuThrSerLeuAsnValSerGluSer 673
352 .....ACTCCTCCAGGTAATGAT..... 369
674 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspSerAlaGlyTh 690
370 .....GAACCTACACT.....GAGCAGA 388
690 rLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThr 707
389 TGGGCAACTGTGGAGAGA.....AGGCTGAAAGGATCCACCTCCTCCT 432
      ||| : : : : : : : : : : : : : : : : : : : : : :
707 hrPheAsn.ValGluArgAsnAlaArgValAsnPheAspIleLysAlaPr 723
433 GATTTCAATTGCAGGAAAAAGTTACTGTAATATGGACCACAGGTAGGCG 482
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
723 oIleglyIleAsnLysTyrSerSerLeuAsnTyrAlaSerPheAsnGlyA 740
483 ATTTCTCCATGAGTGGGCTCATCTACGATGGGAGCTATTTCAGCAGTACA 532
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
740 snIleSerValSerGly.....GlyGlySerValAspPheThr 752
533 ATAATGATGAGAAATCTCTATTATCCAAATGGGAAGATACA.....AGCA 576
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
753 LeuLeuAlaSerSerSerAsn.ValGlnThrProGlyValValIleAsnS 769
577 GTAACATGTTACAGCAGGTATTACTGTTCAAAATGTAAGAACAGTGTCA 626
      ||||| : : : : : : : : : : : : : : : : : : : : : :
769 erLysTyrPheAsn.ValSerThrGlySerLeuArgPheLysThrSe 785
```



```

627 GGA.....GGCAGCTTTACACAAAAGATGCACATTCAT. 663
:|||||
785 rGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeuAsnA 802
:|||||
664 .....AAGTAACAGGACTCTATGAAAAGGATGTGAGTTGTT 702
:|||||
802 laThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspGlyMet 818
:|||||
703 CTCCAATCCGCCAGAGCGGAGGCTTCTATATATGTTT..... 741
:|||||
819 ileGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyGlyAs 835
:|||||
742 .....GCACACATGTTGATTCATATAGTTG 766
:|||||
835 nIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsnValT 852
:|||||
767 AATTCCTGACAGAA.....CAAAC 786
:|||||
852 hrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPheAspAsn 868
:|||||
787 CACAACAAGAAGCTCCAACAAGCAAAAT.....CAAAAATGCAA 827
:|||||
869 HisGlnLysProLeuThrIleLysLysAspValIleIleAsnSerGlyAs 885
:|||||
828 TCTCCGAAGCACATGGGAAGTG.....ATCC 853
:|||||
885 nLeuThrAlaGlyGlyAsnIleValAlaIleAlaGlyAsnLeuThrValG 902
:|||||
854 GTGATTCCTGAGGACTTTAAAGAAACCACCTCTCATGACA..... 891
:|||||
902 luSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsnValGly 918
:|||||
892 .....ACACAGCCACCAATCCACCTCTCATGCTGCAGATGG 932
:|||||
919 GlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLysGlyG 935
:|||||
933 ACAAGAATGTGTGTTAGTCTTGTGACAATCTGGAAGCATGCCGAGTG 982
:|||||
935 yAlaArgPheLysAspIleAspAsnSerLysAsnLeuSerIleThrThra 952
:|||||
983 GTAACCCGCTC.....AATCGACTCAATCAA 1008
:|||||
952 snSerSerSerThrTyrArgThrIleIleSerGlyAsnIleThrAsnLys 968
:|||||
1009 GCAGCCAGCTTTTCCTGTCGACAGAGTTGAGCTGGGGCTCGGTGGTGG 1058
:|||||
969 AsnGlyAspLeuAsnIleThrAsnGlu.....GlySer..... 979
:|||||
1059 GATGTCACATTTGCACAGTGTGCCCATGTACAAGTGAACATCATACAG. 1107
:|||||
980 .....AspThrGluMetGlnIleGlyAspValSerGlnL 992
:|||||
1108 .....ATAAACAGTGGCAGTGACAGGGACACACTGCCCAAAGA 1146
:|||||
992 ysGluGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrLysGln 1008
:|||||
1147 .....TTACCTGCAGAGCTTCAGAGGGAGCTCCATCTGCAGCGGGCT 1190
:|||||
1009 IleThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAspAlaTh 1025
:|||||
1191 TCGATCGCATTTACTGTGATTAGGAAGAA..... 1221
:|||||
1025 rAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnA 1042
:|||||
1222 .....TATCCAACCTGATGGA 1236
:|||||
1042 spLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGly 1058
:|||||
1237 TCTGAATTCGTGCTGACGGATGGGAAGACAACTATAAGTGGG... 1284
:|||||
1059 SerAsp.....LeuThrIleGlyAsnThrAsnSerAlaAspGlyTh 1072
:|||||

```

```

1285 .....TGCTTTAACGAGGTCAAAACAAAGT.....GGTG 1312
:|||||
1072 rAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089
:|||||
1313 CCATCATCCACACAGTCGCTTTGGGCCCTCTGCAGCTCAAGAACTAGAG 1362
:|||||
1089 laAspGlyHisLysValThrLeuHis..... 1097
:|||||
1363 GAGCTGTCCAAATG...ACAGGAGGTTTACAGACATATGCTTCAGATCA 1409
:|||||
1098 .....SerLysValGluThrSerGlySerAsnAsnThrGluAspSe 1112
:|||||
1410 AGTTCAGAACAATGGCCCTCATGTGCTTTGGGCCCTTTTCATCAGGAA 1459
:|||||
1113 rSerAspAsnAla..... 1117
:|||||
1460 ATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGGATTAACC 1509
:|||||
1118 .....GlyLeuThrIleAspAlaLysAsnValThr 1127
:|||||
1310 CTCAGAACAGCCAGTGGATGAATGGCACAGTGTGTCGACAGCACCGT 1559
:|||||
1128 ValAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144
:|||||
1560 GGGAAAGGACACTTTGTTTCTTATCCTCGACACAGCAGCCTCCCCAAA 1609
:|||||
1144 rGlyGlu.....IleThrThrLysThr..... 1151
:|||||
1610 TCCITCTCTGGGATCCCGAGTGGACAGAGCAAGGTGGCTTTGTAGTGGAC 1659
:|||||
1152 .....GlyThrThrIleAsn 1156
:|||||
1660 AAAAACACCAAAATGGCTTACTCTCCAAATCCAGGCATTTGTAAGTTGG 1709
:|||||
1157 AlaThrThrGly.....AsnValGluIleThrAlaGlnThrG 1169
:|||||
1710 CACTTGGAAATACAGTCTGCAAGCAAGCTCACAAACCTTGACCCCTGACTG 1759
:|||||
1169 ySerIleLeuGlyGlyIleGluSerSerSerGlySerValThrLeuThra 1186
:|||||
1760 TCAGTCCCGTGGC.....TCCAATGTACCTGCCCTCCAATTACA 1800
:|||||
1186 laThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
:|||||
1801 GTGACTTCCAAAACAGACAGACACACAGC..... 1830
:|||||
1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLy 1219
:|||||
1830 ..... 1830
:|||||
1219 sGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyGlyT 1236
:|||||
1831 .....AAATTCGCCAGCCCTCTGTGAGTT 1854
:|||||
1236 hrIleSerGlyThrValGluValLysAlaThrGluSerLeuThrThr 1252
:|||||
1855 TATCCAATATTCGC.....CAGGAGCCTCCCAATTCCTCAG 1892
:|||||
1253 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 1269
:|||||
1893 GGCCAGTCTCACAGCCCTGATTGAATCAGTCAATGGAAAAACACTTACCT 1942
:|||||
1269 rAlaThrGlyThr...IleGlyGlyThrIleSerGlyAsnThrValAsnV 1285
:|||||
1943 TGGAA.....CTACTGGATAATGGAGCAGGCTGCTGAT 1974
:|||||
1285 alThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 1301
:|||||
1975 GCTACTAAG.....GATGACGGTGTCTACTCAAG 2003
:|||||
1302 AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 1318
:|||||
2004 GTATTTCAACAACCTTATGACACAGAAATGGTAGATACAGTGTAAAGTCCGGG 2053
:|||||

```

```
1318 rGluAlaSerSerHisIleThrSerAlaLysGlyGlnValAsnLeuSerA 1335
      :::::::::: :::::::::: :::::::::: ::::::::::
2054 CT.....CTGGGAGAGTTAAGCCAGCCAGCGAGAGTG 2088
      || :::::::::: :::::::::: ::::::::::
1335 laGlnAspGlySerValAlaGlySerIleAsnAlaAa.....AsnVal 1349
      :::::::::: :::::::::: ::::::::::
2089 ATACCCAGCAGAGTGGAGCACTGTACATACCTGCTGGATTGAGAATGA 2138
      :: :::::::::: :::::::::: ::::::::::
1350 ThrLeuAsnThrGlyThrLeuThr.....ValLysGlySe 1363
      :: :::::::::: ::::::::::
2139 TGAATACAAATGAATCCACCAGACCTGAATTAATTAAGGATGATGTC 2188
      :::::::::: :::::::::: ::::::::::
1363 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 1380
      :::::::::: ::::::::::
2189 AA.....CACACGAAGTGTGTTTCAGCAGA 2214
      :: ::::::::::
1380 LuLeuAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn 1396
      ::::::::::
2215 ACATCTCCGGAGGCTCATTTCTGGCTTCTGATGTCCCAAAATGCTCCCAT 2264
      :::::::::: ::::::::::
1397 AlaAsnGlySerGlySerValIleAlaThrThrSerArgValAsnIl 1413
      ::::::::::
2265 ACCT...GATCTCTCCACCTGGCCAAATC.....ACCG 2296
      ||:::::::: ::::::::::
1413 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleSerLysAsnG 1430
      ::::::::::
2297 ACTGAAGCGGAAATTCACGGGGCAGCTCATTAATCTGACTTGACCA 2346
      ::::::::::
1430 lYleAsnThrValLeuLeuLysGlyValLysIleAspValLysTyrIle 1446
      ::::::::::
2347 GCTCTGGG...GATGATTATGACCATGGAACAGCTCACAAATATATCAT 2393
      ||:::: ::::::::::
1447 GlnProGlyIleAlaSerValAspGluValIleGluAlaLys..... 1460
      ::::::::::
2394 TCGAATAGTACAGTATTTCTGATCTCAGACAGATTCATGATGATC 2443
      ||:::: ::::::::::
1461 .ArgIleLeuGluLysValLysAspLeuSerAspGluGluArgGluAlaL 1477
      ::::::::::
2444 TT 2445
      ||
1477 eu 1477
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-617-697-10

```
seq_documentation_block:
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-10

alignment_scores:
      Quality: 138.50      Length: 680
      Ratio: 0.425      Gaps: 29
      Percent Similarity: 47.941      Percent Identity: 20.882

alignment_block:
US-09-049-696-18 x US-08-617-697-10 ..

Align seg 1/1 to: US-08-617-697-10 from: 1 to: 1600

      865 GACITTTAAGAAACACACCTATGACAAACACAGCCACCAATCCACCTT 914
      ||| :::::::::: ::::::::::
      86 AspGlyAsnLysThrIleArgAsnSerValAsnAlaIleAsnTr 102
      ||| ::::::::::
      915 CTCATTCGTGCAGATTGCACAAAGAAATTTGTGTTTTCCTTGACAAAT 964
      ||| ::::::::::
      102 physGlnPheAsnIleAspGlnAsnGluMetGluGlnPheLeuGlnGlu 119
      ||| ::::::::::
      965 CTGGAAGCATGCGGACTGGTACCGCCCTCAATCGA..... 999
      ||| ::::::::::
      119 erSerAsnSerAlaValPheAsnArgValThrSerAspGlnIleSerGln 135
      ||| ::::::::::
      1000 .....CTGAATCAAGCAGCGCAGCTTTTCCTGCTGCAG..... 1032
      ||| ::::::::::
      136 LeuLysGlyIleLeuAspSerAsnGlyGlnValPheLeuIleAsnProAs 152
      ||| ::::::::::
      1033 .....ACAGTTGAGCTGGGCTCTGGTTCGGATGGTGACATTTGAC. 1074
      ||| ::::::::::
      152 nglyIleThrIleGlyLysAspAlaIleIleAsnThrAsnGlyPheThrA 169
      ||| ::::::::::
      1075 ..AGTGCTGCCCATGTACAAAGTGAACATACAGATAACAGTGGCAGT 1122
      ||| ::::::::::
      163 laSerThrLeuAspIleSerAsnGluAsnIleLysAlaArgAsnPheThr 185
      ||| ::::::::::
      1123 GACAGGGACACACTCGCCAAAGATTACCTGCAGCAGCTTCAGGAGGGAC 1172
      ||| ::::::::::
      186 LeuGluGlnThrLysAspLysAlaLeuAlaGluIleValAsnHisGlyLe 202
      ||| ::::::::::
      1173 GTCCATCTCCAGCGGCTCGATCGCATTTACTGTGATTAGGAAGAAT 1222
      ||| ::::::::::
      202 u...IleThrValGlyLysAspGlySerValAsnLeuIleGlyLysV 218
      ||| ::::::::::
      1223 ATCCAACATGAT.....GGATCTGAAATTTGCTGCTG 1254
      ||| ::::::::::
      218 aLysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeuLeu 234
      ||| ::::::::::
      1255 ACAGATGGGAGACACAACTACTAGTGGTGTCTTAAACGAGGTCAACA 1304
      ||| ::::::::::
      235 Ala...GlyGlnLysIleThrIleSerAspIleIleAsnPro..... 247
      ||| ::::::::::
      1305 AAGTGTGTCATCATCCACAGCTCCGCTTTGGGCCCTCTCGAGCTCAAG 1354
      ||| ::::::::::
      248 .....ThrIleThrIleAlaAlaProGluAsnGluAlaIleA 262
```

```
1355 AACTAGAGGAGCTCTCCAAATGACAGGAGGTTTACAGACATATGCTTCA 1404
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
262 snLeuGlyAspIlePheAlaLysGlyGlyAsnIleAsnValArgAlaAla 278
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1405 GATCAAGTTTCAGAACAAATGGCTCATT.....GATGCTTTTGGGGCCT 1448
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
279 ...ThrIleArgAsnLysGlyLysLeuSerAlaAspSerValSerLysAs 294
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1449 TTCAATCAGGAATGGAGCTGTCTCTCAGCGCTCCATCCAGCTCAGAGTA 1498
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
294 pLysSerGlyAsnIleValLeuSerAlaLysGlyGlyAlaGluAluIleG 311
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1499 AGGGA...TTAAACCTCCAGAACAGCCAGTGGATGAATGCACAGTGATC 1545
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
311 lyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGlyLysLeuMet 327
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1546 GTG..... 1548
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
328 ileThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLeuSe 344
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1549 .....GACAGCACCGTGG 1561
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
344 rGlyLysGluGlyGlyGluThrTrpLeuGlyGlyAspGluArgGlyGluG 361
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1562 GAAAGGACACTTTCTTCTTATCACCTGGACAGCGAGCCTCCCAAAATC 1611
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
361 lyLysAsnGlyIleGlnLeuAlaLysLysThrLeuGluLysGlySer 377
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1612 CTCTCTCGGATCCAGTGCAGAGCAAGAGTGGC..... 1647
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
378 ThrIle...AsnValSerGlyLysGluLysGlyGlyArgAlaIleValTr 393
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1648 .....TTTGTAGTCGACAAAACACACAAATGGCTACCTCCAAA 1687
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
393 pGlyAspIleAlaLeuIleAspGlyAsnIleAsn.....AlaGlnG 407
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1688 TCCAGCAGTATTGTAAAGTTGGCACTTGGAAATACAGTCTGCAAGCAAG 1737
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
407 lySerAspIleAlaLysThrGlyGlyPhe.....ValGluThrSer 420
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1738 TCACAAACCTTGACCCG..... 1755
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
421 GlyHisAspLeuSerIleGlyAspValIleValAlaLysGluTr 437
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1756 .....ACTGTCAGCTCCCGTCCGT 1774
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
437 pLeuLeuAspProAspAspValSerIleGluThrLeuThrSerGlyArg 454
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1775 CCAATGCTACCCCTCCCTCCAAATACAGTGACTTCCAAAACGACAAAGGAC 1824
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
454 snAsnThrGlyGluAsnGlnGlyThrThrGlyAspGlyThrLysGlu 470
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1825 ACCAGCAAA.....TTCCCCAGCCCTCTGTAGTTATTCACAAATAT 1865
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
471 SerProLysGlyAsnSerIleSerLysProThrLeuThrAsnSerThrLe 487
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1866 TCGCCAGGAGCCCTCCCAATCTCAGGGCAGTGTACAGCCCTGATTG 1915
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
487 uGluGln.....IleLeuArgArgGlySerTyrValAsnIleT 500
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1916 AATCAGTGAATGGAACAAACAGTTACCTTGGAACTA...CTGGATAATGGA 1962
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
500 hrAlaAsnAsnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGly 516
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1963 GCAGGTGCTGATGCTACTAAGGATGACGGTGTCTACTCAAGGATTATTCAC 2012
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
517 SerLeuThrLeuHisThrLysArgAspGlyValLysIleAsnGlyAspIle 533
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2013 AACTTATCAGCAAGTGGTATACAGTGTAAAGTCGGGGCTCTGGAG 2062
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
533 eThrSerAsnGluAsnGlyAsnLeuThrIleLysAlaGlySerTrpVala 550
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2063 GAGTTACGCGACGACGAGGAGTGTATACCCAGCAGAGTGGACACTG 2112
```

```
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
550 spValHis.....LysAsnIleThrLeuGlyThrGlyPheLeu 562
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2113 TACATACCTGGCTGGATTGAGAAATGATAACAATCGAATCCACCAAG 2162
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
563 AsnIle.....ValAlaGlyAspSerValAlaPheGluArgGluI 576
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2163 ACCTGAA..... 2169
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
576 yAspLysAlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrI 593
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2170 .....ATTAATAAGGATGATGTTCAACACACAG.....CAAGTGTGTTTC 2208
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
593 leThrValAsnLysAspLysGlnPheArgPheAsnAsnValSerLeu 609
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2209 AGCAGAACATCCTCGGAGGCTCATTTGTGGCTTCTGATGTCCTCCAAATGC 2258
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
610 AsnGlyThrGlyLysGlyLeuLysPheIleAlaAsnGln..... 622
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2259 TCCCATACCTGATCTCTCCACCTGGCCAAATCACC.....GACC 2299
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
623 .....AsnAsnPheThrHisLysPheAspG 631
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2300 TGAAGCGGAAATTCACGGGGCAGTCTCATTAATCTGACT..... 2340
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
631 lyGluIleAsnIleSerGlyIleValThrIleAsnGlnThrThrLysLys 647
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2341 .....TGGACAGCTCTCGGGCATGATTATGACCAT..... 2370
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
648 AspValLysTyrTrpAsnAlaSerLysAspSerTyrTrpAsnValSerSe 664
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2371 .....GGAACAGCTCACAGTAT.....ATCATTGAA 2398
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
664 rLeuThrLeuAsnThrValGlnLysPheThrPheIleLysPheValAspS 681
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2399 TAAGTACAAGTATTCTTGATCTCAGACAGCAAGTTCAATGATCTCTCAA 2448
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
681 erGlySerAsnGlyGlnAspLeuArgSerSerArgArgSerPheAlaGly 697
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2449 GTGAATACTACTGCTCTCATCCCAAGGAA.....GCCAA 2483
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
698 ValHisPheAsnGlyIleGlyGlyLysThrAsnPheAsnIleGlyAlaAs 714
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2484 CTCTGAGGAAGTCTTTTGTGTTTAAACCAGAAACATTACT 2523
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
714 nAlaLysAlaLeuPheLysLeuLysProAsnAlaAlaThr 727
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
seq_name: /cqn2_6/plodata/2/1aa/6B_COMB.pep:US-08-296-791-6
seq_documentation_block:
; Sequence 6, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4,Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
```

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RET/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; US-08-296-791-6

alignment_scores:
  Quality: 130.00      Length: 956
  Ratio: 0.332        Gaps: 33
  Percent Similarity: 41.004      Percent Identity: 21.234

alignment_block:
US-09-049-696-18 x US-08-296-791-6 ..
Align seg 1/1 to: US-08-296-791-6 from: 1 to: 1848

349 TCTACTCTCCAGGT.....AATGATCAACCTACACTGACAG..... 387
||| ||| ||| :|||:||||| |||:|
278 SerAspProLysGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaVa 294
388 .ATGGCAACTGTGA.....GAGAAGGTG 412
:|:|:|:|:|:| ||| |||:|:|:|:|:|
294 lleuGlySerGlySerProLeuPheValTyrAspArgGluLysGlyL 311
||| ||| ||| :|||:||||| |||:|
413 AAAGATCCACCTC...ACTCTGATTTTCATTCGACGAGAAAGTTAGCT 459
:|:|:|:|:|:| ||| ||| |||:|:|:|:|
311 ysTrpLeuPheLeuGlySerTyrAspPheTrpAlaGlyTyrAsnLys 327
460 GAATATGGACCAAGGATGGCATTTCATGATGAGTGGGCTCATCTACG 509
:|:|:|:|:|:| ||| ||| |||:|:|:|:|
328 SerTrp...GlnGluTrpAsnIleTyrLysHisGluPheAla..... 340
510 ATGGGGAGTATTGTCAGGAGTACAATAATGATGAGAAATTCATTATCCA 559
||| ||| ||| :|||:||||| |||:|
341 .....GluLysIleTyr..... 344
560 ATGGAAGATAACAGCAGTAAGATGTTCCAGAGGT...ATTACTGTGACA 606
||| ||| ||| :|||:||||| |||:|
345 .....GlnGlnTyrSerAlaGlySerLeuIleGlySer 355
607 AATGTAGTAAGAAGTGTACGGGAGGAGCTTTACACCAAAAGATGCAC 656
||| ||| ||| :|||:||||| |||:|
356 Asn.....ThrGlnTyrThrTrpGlnAlaTh 364
657 A.....TTCAATAAAGTAACAGGACTCTCTGATAAAGGATGTGAGTTG 700
||| ||| ||| :|||:||||| |||:|
364 rGlySerThrThrIleThrGly..... 372
701 TTCTCCAATCCCGCAGGAGAGAGGTTCTATATGTTTCACAACAT 750
||| ||| ||| :|||:||||| |||:|
373 .....GlyGlyGluProLeuSerValAspLeuThrAspGly 384
751 GTTGATTCTATAGTTGAATTCGTACAGACAAACCAACCAACAAAGC 800
||| ||| ||| :|||:||||| |||:|
385 LysAsp.....LysProAsnHisGlyLysSerI 394
801 TCCAACCAAGCAAAATCAAAAATGCAATCTCCGA..... 834
||| ||| ||| :|||:||||| |||:|
394 eThrLeuLysGlySerGlyThrLeuThrLeuAsnHisIleAspGlnG 411
834 ..... 834

```

```

411 lyAlaGlyGlyLeuPheGluGlyAspTyrGluValLysGlyThrSer 427
835 .....AGCACATGGGAAGTATCCCTGATTCCTGAGGACTTTAAGAAAC 878
:|:|:|:|:|:| ||| |||:|:|:|:|
428 AspSerThrThrTrpLysGlyAlaGlyValSerValAlaAspGlyLysTh 444
879 CACTCTATGACACACAGCCACCAAAATCCACCTTCTCA...TTGCTGC 925
:|:|:|:|:|:| ||| ||| |||:|:|:|:|
444 r.....ValThrTrpLysValHisAsnProLysTyrAspArgLeuAlaL 459
926 .AGATTGGACAAAGAATTGCTGTTTAGTCCTTGAC.....AAA 963
:|:|:|:|:|:| ||| ||| |||:|:|:|:|
459 ysIleGlyLysGly.....ThrLeuValGluGlyLysGlyLysAsn 473
964 TCTGGAAGCATGGCGACTGGTAACCGCCTCAATCGACTCAATCAAGCAGG 1013
:|:|:|:|:|:| ||| ||| |||:|:|:|:|
474 GluGlyLeuLeuLysValGlyAsp.....G 482
1014 CCAGCTTTTCTGCTGCAG.....ACAGTTGAGCTGG 1045
||| ||| ||| :|||:||||| |||:|
482 yThrValIleLeuLysGlnLysAlaAspAlaAsnLysValGlnAlap 499
1046 GGTCTGCTGGTGGGATGGTACATTTGACAGTGTGCTGCCCATGTACAAAGT 1095
||| ||| ||| |||:|:|:|:| |||:|
499 heSerGlnValGlyIleValSerGlyArgSerThr..... 510
1096 GAACATACACAGATAACAGTGGCAGTGCACAGGACACACTGCCCAAAG 1145
:|:|:|:|:|:| ||| ||| |||:|:|:|:|
511 .....LeuValLeuAsn.....AspAspLysGlnValAspProAsnSe 523
1146 ATTACCTGCAGCAGCTTCAGGAGGACGCTCCATCTGCAGCGGCTTCGAT 1195
:|:|:|:|:|:| ||| ||| |||:|:|:|:|
523 rIleTyrPheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnSerL 540
1196 CGGCTTTTACTGTATTAGGAAGAAATATCCAACTGATGGATCTGAAATT 1245
:|:|:|:|:|:| ||| ||| |||:|:|:|:|
540 euThrPheAspHisIleArgAsn.....IleAspAspGlyAlaArgVal 554
1246 GTG.....CTGCTCAGCGATGGGGAAGACACACTATAAGTGG..... 1284
||| ||| ||| :|||:||||| |||:|
555 ValAsnHisAsnMetThrAsnThrSerAsnIleThrIleThrGlyGluSe 571
1284 ..... 1284
571 rLeuIleThrAsnProAsnThrIleThrSerTyrAsnIleGluAlaGlnA 588
1285 .....TGC 1287
588 spAspAspHisProLeuArgIleArgSerIleProTyrArgGlnLeuTyr 604
1288 TTTAACGAGGTCAACAAGTGGTCCCATCATCCACACAGTCGCTTGGG 1337
||| ||| ||| :|||:||||| |||:|
605 PheAsnGlnAspAsnArgSer.....TyrTyrThrLeuLysLysG 618
1338 GCGCTCTGCAGCTCAAGAACTAGAGAGCTGTCCAAAATGACA..... 1380
||| ||| ||| :|||:||||| |||:|
618 yAlaSerThrArgSerGluLeuProGlnAsnSerGlyGluSerAsnGluA 635
1381 .....GGAGGTTTACAGACATATGCTTCAGATCAAGTTGAG 1416
||| ||| ||| :|||:||||| |||:|
635 snTrpLeuTyrMetGlyArgThrSerAspAlaAlaLysArgAsnValMet 651
1417 AAC.....AATGGCTCATTTGATCTTTTGGGCG 1445
||| ||| ||| :|||:||||| |||:|
652 AsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrPheGly... 667
1446 CTTTTTCAGGAAATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGA 1495
:|:|:|:|:|:| ||| ||| |||:|:|:|:|
668 .....GluGluGluThrLysAlaThrGlnAsnGlyLysLeuAsnValThrP 683
1496 GTAAGGGATTAACTCCAGACACACCCAGTGCATCAATGGCACACTGATC 1545
||| ||| ||| :|||:||||| |||:|
683 heAsnGlyLysSerAspGlnAsnArgPheLeuLeuThrGlyGlyThrAsn 699

```

[illegible][illegible]

This Page Blank (uspto)

OM of: US-09-049-696-18 to: A_Geneseq_1101.* out_format : pfs

Date: Mar 30, 2002 2:19 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framet_n2p.model -DEV=xlip
-Q=/cgn2_1/USPTO_spool/US09049696/runat_28032002_145237_2001/app_query.fasta_1.12579
-DB=A_Geneseq_1101 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09049696_@CGN1_1_293 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-18

Query length: 2813

Database: A_Geneseq_1101.*

Database sequences: 522463

Database length: 74073290

Search time (sec): 819.910000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SID52/gcgdata/geneseq/geneseq/AA1999.DAT:AA11789 +		199.50	321.62	92	6.6e-10
/SID52/gcgdata/geneseq/geneseq/AA2000.DAT:AA03643 +		193.50	311.78	91	2.4e-09
/SID52/gcgdata/geneseq/geneseq/AA1997.DAT:AAW30292 +		143.50	193.26	1601	0.0005
/SID52/gcgdata/geneseq/geneseq/AA2000.DAT:AAW30292 +		143.50	193.26	1601	0.0005
/SID52/gcgdata/geneseq/geneseq/AA1996.DAT:AAW9393 +		141.50	187.37	2411	0.0010
/SID52/gcgdata/geneseq/geneseq/AA1993.DAT:AAW41732 +		140.50	188.84	1529	0.0010
/SID52/gcgdata/geneseq/geneseq/AA1993.DAT:AAW41725 +		139.50	187.13	1536	0.0012
/SID52/gcgdata/geneseq/geneseq/AA1998.DAT:AAW56322 +		139.50	183.62	2039	0.0014
/SID52/gcgdata/geneseq/geneseq/AA2000.DAT:AAW1847 +		138.50	189.65	1095	0.0012
/SID52/gcgdata/geneseq/geneseq/AA1993.DAT:AAW41723 +		138.50	185.46	1536	0.0015
/SID52/gcgdata/geneseq/geneseq/AA1994.DAT:AAW63505 +		138.50	185.46	1536	0.0015
/SID52/gcgdata/geneseq/geneseq/AA2000.DAT:AAW01846 +		138.50	185.46	1536	0.0015
/SID52/gcgdata/geneseq/geneseq/AA2001.DAT:AAW89860 +		135.50	184.03	1152	0.0024
/SID52/gcgdata/geneseq/geneseq/AA2001.DAT:AAW79787 +		135.50	184.03	1152	0.0024
/SID52/gcgdata/geneseq/geneseq/AA2001.DAT:AAU05396 +		134.00	142.59	26926	0.0211
/SID52/gcgdata/geneseq/geneseq/AA2001.DAT:AAW98842 +		133.50	176.89	1569	0.0044
/SID52/gcgdata/geneseq/geneseq/AA2000.DAT:AAW01834 +		133.00	180.44	1101	0.0040
/SID52/gcgdata/geneseq/geneseq/AA2001.DAT:AAE01422 +		132.00	179.72	1020	0.0048
/SID52/gcgdata/geneseq/geneseq/AA2001.DAT:AAW31889 +		132.00	161.68	4393	0.0112
/SID52/gcgdata/geneseq/geneseq/AA1997.DAT:AAW30293 +		131.50	173.83	1536	0.0067

seq_name: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT:AA11789

seq_documentation_block:

ID AA11789 standard; Protein: 92 AA.

XX AC AA11789;

XX AC AA11789;

DT 18-JUN-1999 (first entry)

XX Human 5' EST secreted protein SEQ ID No: 389.

DE Human; secreted protein; EST; expressed sequence tag; diagnosis;

KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;

KW upstream regulatory sequence; cytokine activity; cell proliferation;

KW differentiation; haematopoiesis regulation; tissue growth regulation;

KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;

KW thrombolytic; anti-inflammatory; tumour inhibition.

XX Homo sapiens.

OS Homo sapiens.

XX

FW WC3906550-A2.

XX

PD 11-FEB-1999.

XX

PF 31-JUL-1998; 98WO-IB01232.

XX

XX 01-AUG-1997; 97US-0905144.

PR

FA (GEST) GENSET.

XX

XX Duclert A, Dumas Milne Edwards J, Lacroix B;

XX

XX WPI; 1999-153780/13.

DR N-PSDB; AAX40511.

XX

PT New isolated prostate-derived nucleic acids - used to develop

PT products which may have cytokine, immune regulatory, haematopoiesis

PT regulating, anti-inflammatory or tumour inhibition activity

XX

PS Claim 34; Page 546; 675pp; English.

XX

CC AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for

CC human secreted proteins expressed in prostate, and encode the proteins

CC given in AAY11716 to AAY11993 respectively. The proteins given represent

CC the signal peptide and an N-terminal fragment of a secreted protein. The

CC nucleic acid sequences can be used for producing secreted human gene

CC products. They can also be used to develop products for diagnosis and

CC therapy. The proteins obtained may have cytokine activity, cell

CC proliferation and differentiation activity, haematopoiesis regulating

CC activity, tissue growth regulating activity, reproductive hormone

CC regulating activity, chemotactic/chemokinetic activity, haemostatic and

CC thrombolytic activity, receptor/ligand activity, anti-inflammatory

CC activity, tumour inhibition activity or other activities. The products

CC can be used in forensic, gene therapy and chromosome mapping procedures.

CC The sequences can also be used for obtaining corresponding promoter

CC sequences. The nucleic acids encoding the signal peptides can be used for

CC directing extracellular secretion of a polypeptide or the insertion of a

CC polypeptide into a membrane, or importing a polypeptide into a cell.

XX

SQ Sequence 92 AA;

alignment_scores:

Quality: 199.50 Length: 95

Ratio: 2.733 Gaps: 3

Percent Similarity: 76.842 Percent Identity: 45.263

alignment_block:

US-09-049-696-18 x AAY11789 ..

Align seg 1/1 to: AAY11789 from: 1 to: 92

4 ATCACAGGAGATGTACACAAATGGGGCCATTTAAGAGTTCTGTTCAT 53

1 MetThrGlnArgSerIleAla...GlyProIleCysAsnLeuLysPheVa 16

54 CTTGATCTTCACCTTCTAGAGGGGGCCCTGAGTAAATTCACCTC..... 96

16 lThrLeuLeu.....ValAlaLeuSerSerGluLeuProPheL 29

97ATTACGCTGACACAAATGGCGTATGAAGCATTCGCTT 135

29 euGlyAlaGlyValGlnLeuGlnAsnGlyTyrAsnGlyLeuLeu 45

135 GAATCGACCCCAATGCGCAGAGATGAACACTCATTCAACAAATAA 185

45 AlaIleAsnProGlnValProGluAsnGlnAsnLeuLeuSerAsnIle 62

186 GGACATGGTGACCCAGCATCTCTGATCTGTTTGAAGCTACAGGAAGC 235

62 sGluMetIleThrGluAlaSerPheTyrLeuPheAsnAlaThrLysArg 79

236 GATTTTATTCAAAATGTTGCCATTGATTCCT 270
 || :|||:||||: |||:|||||
 79 rgvaipheargAsnIleLysIleLeuIlePro 90

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT: AAG03643

seq_documentation_block:

ID AAG03643 standard; Protein: 91 AA.

XX AAG03643;

XX 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 7724.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.

XX Homo sapiens.

XX EF1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX N-PSDB; AAC03649.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX Claim 13; SEQ ID 7724; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 91 AA;

alignment_scores:

Quality: 193.50 Length: 94
 Ratio: 2.688 Gaps: 3
 Percent Similarity: 76.596 Percent Identity: 44.681

alignment_block:

US-09-049-696-18 x AAG03643 ..

Align seg 1/1 to: AAG03643 from: 1 to: 91

4 ATCAGGGAGATGTACAGCAATGGGCCATTTAAGACTTCTGTTTCAT 53

1 MetThrGlnArgSerIleAla...GlyProIleCysAsnLeuLysPheVa 16

54 CTTGATTCTTCACCTCTAGAGAGGGGCCCTGAGT.....A 88

16 lThr.....LeuLeu**Gly**Ser***LeuPro***LeuG 30
 89 ATTCACTCATTCAGCTGACACAAATGCTATGAGGCAATCTCTCGTTGCA 138
 30 IyAlaGlyValGlnLeuGlnAsnGlyTyrAsnGlyLeuLeuIleAla 46
 139 ATCGACCCCAATGTGCCAGAAGATGAACACTCATTCACAAATAAAGA 188
 47 IleAsnProGlnValProGluAsnGlnAsnLeuIleSerAsnIleLysGI 53
 189 CATGTGACCCAGGCACTCTGTATCTCTTTGAAGCTACAGGAAGCGAT 238
 63 uMetIleThrGluAlaSerPheTyrLeuPheAsnAlaThrLysArgArgV 80
 239 TTTATTTCAAAATGTCGCCATTGATTCCT 270
 80 alPhePheArgAsnIleLysIleLeuIlePro 90

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT: AAW30292

seq_documentation_block:

ID AAW30292 standard; Protein: 1601 AA.

XX AAW30292;

XX 14-APR-1998 (first entry)

XX Non-typeable Haemophilus high mol.wt. surface protein HMW4.

XX Non-typeable Haemophilus; high molecular weight surface protein;

KW HMW4; immunogen; vaccine; otitis media.

XX Haemophilus influenzae strain 5.

XX Key Location/Qualifiers

FT Misc-difference 372 /note= "encoded by TCT"

FT Misc-difference 400 /note= "encoded by AAT"

XX WC9736914-A1.

XX 09-OCT-1997.

XX 01-APR-1997; 97WO-US04707.

XX 01-APR-1996; 96US-0617697.

XX (BARE/) BARENKAMP S J.

XX Barenkamp SJ;

XX WPI; 1997-503038/46.

XX N-PSDB; AAT90993.

XX High molecular weight proteins of non-typeable Haemophilus
 PT Influenzae - useful for vaccine production

XX Claim 1; Page 97-102; 183pp; English.

XX This protein comprises the high molecular weight surface protein
 CC HMW4 (123 kDa) of non-typeable Haemophilus influenzae strain 5 that
 CC has the immunological ability to protect against disease caused by
 CC a non-typeable Haemophilus strain and is characterised by at least
 CC one surface-exposed B-cell epitope that is recognised by monoclonal
 CC antibody AD6. The HMW4 amino acid sequence was deduced from an
 CC isolated hmw4 gene (see AAT90993). HMW1 (see AAW30293), HMW2 (see
 CC AAW30294) and HMW3 (see AAW30291) have also been identified. A
 CC conjugate comprising HMW4 linked to an antigen, hapten or
 CC polysaccharide, and a synthetic peptide of 6-150 amino acids
 CC corresponding to at least protective epitope of HMW4 are also
 CC claimed. HMW proteins, conjugates and peptides can be used in

CC vaccines, as immunogens for preparation of antibodies and as
 CC antigens for detection of these antibodies.

XX
 SQ Sequence 1601 AA;

alignment_scores:

Quality: 143.50 Length: 680
 Ratio: 0.440 Gaps: 29
 Percent Similarity: 47.941 Percent Identity: 20.882

alignment block:

US-09-049-696-18 x AAW30292 ..

Align seg 1/1 to: AAW30292 from: 1 to: 1601

```

865 GACTTTAAGAAAACCACTCTCTATGACACACACAGCCACCAATCCACCTT 914
      ||| :|||:|||||
86 AspGlyAsnLysThrThrIleArgAsnSerValAsnAlaIleIleAsnTr 102
      ||| :|||:|||||
915 CTCAATGCTGCAGATTGGCAAGAATTTGGTCTTTAGTCTTGCACAAAT 964
      ||| :|||:|||||
102 pLysGlnPheAsnIleAspGlnAsnGluMetValGlnPheLeuGlnGluS 119
      ||| :|||:|||||
965 CTGGAAGCATGGCGACTGTTAAACCGCTCAATCGA..... 999
      ||| :|||:|||||
119 erSerAsnSerAlaValPheAsnArgValThrSerAspGlnIleSerGln 135
      ||| :|||:|||||
1000 .....CTGAATCAAGCAGCGCCAGCTTTTCTGCTGCAG..... 1032
      ||| :|||:|||||
136 LeuLysGlyIleLeuAspSerAsnGlyGlnValPheLeuIleAsnProAs 152
      ||| :|||:|||||
1033 .....ACAGTTAGCTGGGTCCTGGTGGGATGGTGACATTTCAC. 1074
      ||| :|||:|||||
152 nGlyIleThrIleGlyLysAspAlaIleIleAsnThrAsnGlyPheThra 169
      ||| :|||:|||||
1075 ..AGTGCTGCCCTGACAAAGTACTCATACAGATAAAGACAGTGCAGT 1122
      ||| :|||:|||||
169 laSerThrLeuAspIleSerAsnGluAsnIleLysAlaArgAsnPheThr 185
      ||| :|||:|||||
1123 GACAGGACACACTCCCAAGAAATTACCTGCAGCAGCTTCAGAGGGAC 1172
      ||| :|||:|||||
186 LeuGluGlnThrLysAspLysAlaLeuAlaGluIleValAsnHisGlyLe 202
      ||| :|||:|||||
1173 GTCCATCTGCAGCGGCTTCGATCGCATTTACTGTGATTAGGAAGAAAT 1222
      ||| :|||:|||||
202 u....IleThrValGlyLysAspGlySerValAsnLeuIleGlyLysV 218
      ||| :|||:|||||
1223 ATCCAACTGAT.....GGATCTGAATTTGGTGGCTG 1254
      ||| :|||:|||||
218 allLysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeuLeu 234
      ||| :|||:|||||
1255 ACGGATGGGAAGACAACTATAAGTGGTGCTTTAACAGAGGTCAAACA 1304
      ||| :|||:|||||
235 Ala...GlyGlnLysIleThrIleSerAspIleIleAsnPro..... 247
      ||| :|||:|||||
1305 AAGTGGTGCCATCATCCACACACTCTCTGGGGCCCTCTGCAGCTCAAG 1354
      ||| :|||:|||||
248 .....ThrIleThrTySerIleAlaLaProGluAsnGluAlaIleA 262
      ||| :|||:|||||
1355 AACTAGGAGAGCTGTCCTCAAAATGACAGGAGGTTTACAGACATATGCTCA 1404
      ||| :|||:|||||
262 snLeuGlyAspIlePheAlaLysGlyLysAsnIleAsnValArgAlaAla 278
      ||| :|||:|||||
1405 GATCAAGTTTCAGAACAAATGCCCTCATT.....GATGCTTTTGGGCCCT 1448
      ||| :|||:|||||
279 ...ThrIleArgAsnLysGlyLysLeuSerAlaAspSerValSerLysAs 294
      ||| :|||:|||||
1449 TTCATCAGAAATGGAGCTCTCTCTCAGCGCTCCATCCAGCTTGAGAGTA 1498
      ||| :|||:|||||
294 pLysSerGlyAsnIleValLeuSerAlaLysGluGlyGluAlaGluIleG 311
      ||| :|||:|||||
1499 AGGGA...TTAACCTCCAGACAGCAGTGGATGAATGGCAGCTGATC 1545
      ||| :|||:|||||

```

```

      ||| :|||:|||||
311 yGlyValIleSerAlaGlnAsnGlnAlaLysGlyGlyLysLeuMet 327
      ||| :|||:|||||
1546 GTG..... 1548
      ||| :|||:|||||
328 IleThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLeuSe 344
      ||| :|||:|||||
1549 .....GACAGACCGCTGG 1561
      ||| :|||:|||||
344 rGlyLysGluGlyGlyGluThrTyThrLeuGlyGlyAspGluArgGlyGluG 361
      ||| :|||:|||||
1562 GAAAGGACACTTTGTTCTTATCAGCTGGACACGAGCGCTCCCAATC 1611
      ||| :|||:|||||
361 yLysAsnGlyIleGlnLeuAlaLysLysThrThrLeuGluLysGlySer 377
      ||| :|||:|||||
1612 CTTCCTCGGATCCAGTCGACAGAACAGGTGGC..... 1647
      ||| :|||:|||||
378 ThrIle...AsnValSerGlyLysGluLysGlyGlyArgAlaIleValTr 393
      ||| :|||:|||||
1648 .....TTGTAGTGCACAAAACACCAAAATGGCGCTACTCCAAA 1687
      ||| :|||:|||||
393 pGlyAspIleAlaLeuIleAspGlyAsnIleAsn.....AlaGlnG 407
      ||| :|||:|||||
1688 TCCAGGCATTGCTAAGTTGGCAGTTGGAATACAGTCTGCAAGCAAGC 1737
      ||| :|||:|||||
407 lySerAspIleAlaLysThrGlyGlyPhe.....ValGluThrSer 420
      ||| :|||:|||||
1738 TCACAAACCTTGACCCCTG..... 1755
      ||| :|||:|||||
421 GlyHisAspLeuSerIleGlyAspValIleValAspAlaLysGluTr 437
      ||| :|||:|||||
1756 .....ACTGTCAGCTCCCGTCGCT 1774
      ||| :|||:|||||
437 pLeuLeuAspProAspAspValSerIleGluThrLeuThrSerGlyArgA 454
      ||| :|||:|||||
1775 CCAATCTACCTCCCTCCCAATTTACAGTGACTTCCAAAAGAACAGGAC 1824
      ||| :|||:|||||
454 snAsnThrGlyGluAsnGlnGlyTyThrThrGlyAspGlyThrLysGlu 470
      ||| :|||:|||||
1825 ACCAGCAAA.....TTCGCCAGCCCTCTGGTAGTTTATGCAAAATAT 1865
      ||| :|||:|||||
471 SerProLysGlyAsnSerIleSerLysProThrLeuThrAsnSerThrLe 487
      ||| :|||:|||||
1865 TCCCAAGGAGCCCTCCCAATTTCTAGGCGCAGTGTACAGCCCTGATTG 1915
      ||| :|||:|||||
487 uGluGln.....IleLeuArgArgGlySerTyValAsnIleT 500
      ||| :|||:|||||
1916 AATCAGTGAATGGAACACAGTTTACCTTGGAACTA..CTGGATAATGGA 1962
      ||| :|||:|||||
500 hrAlaAsnAsnArgIleTyValAsnSerSerIleAsnLeuSerAsnGly 516
      ||| :|||:|||||
1963 GCAGGTGCTGATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCAC 2012
      ||| :|||:|||||
517 SerLeuThrLeuHisThrLysArgAspGlyValLysIleAsnGlyAspI 533
      ||| :|||:|||||
2013 AACTTATGACACGAATGGTAGATACAGTGTAAAAGTCCGGGCTCTGGAG 2062
      ||| :|||:|||||
533 eThrSerAsnGluAsnGlyAsnLeuThrIleLysAlaGlySerTrpValA 550
      ||| :|||:|||||
2063 GAGTTAACGCGACGACGAGAGTAGTATACCCAGCAGAGTGGAGCAGCTG 2112
      ||| :|||:|||||
550 spValHis.....LysAsnIleThrLeuGlyThrGlyPheLeu 562
      ||| :|||:|||||
2113 TACATACCTCGCTGATTCAGAAATGATAATACAATGGAATCCCAACAG 2162
      ||| :|||:|||||
563 AsnIle.....ValAlaGlyAspSerValAlaPheGluArgGluG 576
      ||| :|||:|||||
2163 ACCTGAA..... 2169
      ||| :|||:|||||
576 yAspLysAlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrI 593
      ||| :|||:|||||
2170 .....ATTATAGGATGATGTTCAACACAG.....CAAGTGTGTTTC 2208
      ||| :|||:|||||

```



```

1084  ....AsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsn 1098
2332 AATCTGACTTGACAGCTCTCTGGGGATGATTATGACCATGGA 2373
1099 AsnLeuSerTrpThrAlaLysAlaAspLysTyraAspGly 1112

```

seq_name: /STDS2/gcgdata/geneseq/geneseq/AA1996.DAT.AAR99393

seq_documentation_block:

ID AAR99393 standard; Protein; 2353 AA.

AC AAR99393;

DT 15-JAN-1997 (first entry)

XX Haemophilus adhesion protein HA2.

KW Haemophilus adhesion protein; HA2; hsf protein; vaccine.

XX Haemophilus influenzae type b strain C54.

OS WO9630519-A1.

PN 03-OCT-1996.

XX 22-MAR-1996; 96WO-US04031.

XX 24-MAR-1995; 95US-040995.

PA (UYSL-) UNIV ST LOUIS.

PA (UNIW) UNIV WASHINGTON.

XX Barenkamp SJ, St Geme JW;

XX WPI; 1996-455364/45.

DR N-PSDB; AAT41476.

XX Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in
PT vaccines against H. influenzae infection.

XX Claim 5; Page 66-73; 120pp; English.

XX Haemophilus adhesion protein HA2 (AAR99393) is associated with the
CC formation of surface fibrils involved in adhesion to various host
CC cells; it is also referred to hsf (Haemophilus surface fibrils).
CC Its amino acid sequence was deduced from a genomic DNA clone
CC (AAT41476) derived from Haemophilus influenzae type b strain C65.
CC Large quantities of recombinant HA2 can be produced in transformed
CC prokaryotic or eukaryotic host cells, for use in vaccines against
CC H. influenzae infection.

XX Sequence 2353 AA;

```

alignment_scores:
  Quality: 141.50      Length: 967
  Ratio: 0.321        Gaps: 44
  Percent Similarity: 45.605      Percent Identity: 18.718

```

alignment_block:

US-09-049-696-18 x AAR99393 ..

Align seg 1/1 to: AAR99393 from: 1 to: 2353

```

274 ACATGGAGACAAAGCGTACTATGTGACACAAACCTGAGACTACAA 323
    ::::|||||:
429 SerTrpLysAlaLysAlaGlu..... 435
324 AAATGCTGATGTTCTGGTGTGAGTCTCTCCAGGTAATGATGAAC 373
    |||||
436 ....AlaAsp.....ThrAspGlyAlaLeuGluGlyT 445

```

```

374 CCTACACTGACGACATGGCGCAACTGTGGACAGAGGCTGAAGATCCAC 423
    ::::|||||
445 leSerLysAspGlnGluValLysAlaGlyGlu..... 455
424 CTCACTCTCTGATTTTCATTGTCAGGAAAAAAGTTAGCTGAATATGACCACA 473
    :|||:|||||:|||||:
456 ..ThrValThrPheLysAlaGlyLysAsnLeu...LysValLysGlnAs 470
474 AGGTAGGGCATTTTGTCCATGAG.....TGGGCTC 502
    :|||:|||||:
470 pGlyAlaAsnPheThrTySerLeuGlnAspAlaLeuThrGlyLeuThrs 487
503 ATCTACGATGGGAGATATTTGACGAGTACAAATATGATGAGAAATTTCTAC 552
    :|||:|||||:
487.erlleThrLeuGlyGlyThrAsnGlyGlyAsnAspAlaLysThrVal 503
553 TTATCC...AATGGAGAATACAGCAGTAAGATGTTTCAGCAGGTATAC 599
    :|||:|||||:
504 lleAsnLysAspGlyLeuThrIleThrProLagLysnGlyGlyThrTh 520
600 TGGTACAAATGTAGTAAAGTGTCTAGGAGGAGCAGCTGTTACACCAAAA 649
    :|||||:|||||:
520 rGlyThrAsnThrIleSerValThrLysAspGlyIleLysAlaGlyAsnL 537
650 GATGCACATTTCAATAAAGTAAACAGGACTC.....TATGAAAAAGGATGT 693
    :|||:|||||:
537 ysAlaIleThrAsnValAlaSerGlyLeuArgAlaTyraAspAlaAsn 553
694 GAGTTTGTCTCCAATCCCGCAGCAGGAGGAGGCTTCTATAATGTTTGC 743
    :|||||:
554 PheAspValLeuAsnAsnSerAlaThrAsp.....LeuAs 565
744 ACAACATGTTGATTTCTATAGTTGAATTTCTGTACAGAACAAACACACA 793
    :|||||:
565 nArgHisValGluAspAlaTyrlLysGlyLeuLeu.....AsnLeuAsnG 580
794 AAGAAAGTCCCAACCAAGCAAAATCAAAATCAATCTCCGAAGCACATGG 843
    :|||||:
580 luLysAsnAlaAsnLysGln..... 586
844 GAAGTGATCCGTGATCTGTGAGGACTTTAAGAAAACCACTCTATGCAAC 893
    :|||||:
587 .....ProLeuValTh 590
894 ACAGCCACCAATCCCACTCTCATTTGTCGACAGATTGGACAAGAATTG 943
    :|||:|||||:
590 rAspSerThAlaAlaThrValGlyAspLeu.....ArgLysLeuG 604
944.TGTGTTTGTAGTCTTGACAAATCTGGAAGCATGGCGAGTGGTAAACCGCTC 993
    :|||:|||||:
604.lyTrpValValSerThrLysAsnGly.....ThrLysGluGluSer 617
994 AATCGACTGAATCAAGCAGCGCAGCTTTTCTGCTGCAGACAGTTGAGCT 1043
    :|||||:
618 AsnGlnValLysGlnAlaAspGluVal.....Le 627
1044.GGGGTCTCTGGTGGATGTTGACATTTGACAGTGCTGCCCATGTACAAA 1093
    :|||:|||||:
627 uPheThrGlyAlaGlyAlaAlaThrValThrSerLysSerGluAsnGlyL 644
1094 GTGAACATCAGATACAGATAAACAGTGGC..... 1119
    :|||:|||||:
644 ysHisThrIleThrValSerValAlaGluThrLysAlaAspCysGlyLeu 660
1120...AGTGACAGGGACACACTCGCCAAAAGATTACCT..... 1152
    :|||||:
661 GlulysAspGlyAspThrIleLysLeuLysValAspAsnGlnAsnThras 677
1153.....GCAGCAGCTTCAGGAGGAGCTCCATCTGTCAGCGGGCTTC 1192
677 pasnValLeuThrValGlyAsnAsnGlyThrAlaValThrLysGly.... 692
1193 GATCGGCATTTACTGTGATTAGGAAGAAATATCAACTGATGATCGAA 1242

```

```

693 .....GlyPheGluThrVal...LysThrGlyAlaThrAspAlaAspArg 706
1243 ATTGTGCTGTCGCGGATGGGAGACACACTATAAGTGGGTCTTAA 1292
707 .....GlyLysValThrValLysAspAlaThrAlaAs 717
1293 CGAGGTCAACAAGTGGTCCCATCATCCACAGCTCGCTTTGGGCCCT 1342
717 nAspAlaAspLysLysValAlaThrValLysAspValAlaThrAlaAla 734
1343 CTGCAGCTCAAGAACTAGACAGCTGTCCAAATAGCAGAGGTTTACAG 1392
734 snSerAlaAlaThrPheValLysThrGluAsnLeuThrThrSerIleAsp 750
1393 ACATATGCTTCAGATCAAGTTCAGACAAATGGCTCATGATGCTTTGG 1442
751 .....GluAspAsnProThrAspAsnGlyLysAspAsp..... 761
1443 GGCCCTTTCATCAGGAAT.....GGAGCTGTCTCTCAGCGCT 1480
762 .AlaLeuLysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuL 778
1481 CCATCCAGCTTGAGTAAGGATTAAACCTC.....CAGAACAGC 1521
778 ysValLysArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeu 794
1522 CAGTGGATGAATGCCAGAGTGTGGAGCAGACCCGTGGGAAGGACAC 1571
795 GluValLysThrAlaLysValSerAspThrLeuThrIleGlyGlyAsnTh 811
1572 TTTGTTTCTTATCACCTGGACAGCAGCTCCCAATC..... 1611
811 r.....ProThrGlyGlyThrThrAlaThrProLysValAsnIleThrs 826
1611 ..... 1611
826 erThrAlaAspGlyLeuAsnPheAlaLysGluThrAlaAspAlaSerGly 842
1612 .....CTTCTCTGGATCCCG 1628
843 SerLysAsnValTyrLeuLysGlyIleAlaThrThrLeuThrGluProSe 859
1629 TGCACAGAACCAAGCTCGCTTTGTA.....GTGGACAAAAACACCA 1669
859 rAlaGlyAlaLysSerHisValAspLeuAsnValAspAlaThrLysL 876
1670 AAATGGCCTACCTCCAAATCCAGGATGCTAAGGTGGCACTTGGA 1719
876 ysSerAsnAlaAlaSerIleGluAspValLeuArgAlaGly..... 889
1720 TACAGTCTGCAAGCAAGCTCACAACTTGACCTGACTGTACGTCGCCG 1769
890 TrpAsnIleGlnGlyAsnGlyAsnValAspTyrValAlaThrTyrAs 906
1770 TGGCTCCAATGCTACC.....CTGCCTCCCAATTACAGTGA 1804
906 pThrValAsnPheThrAspAspSerThrGlyThrThrThrValThrVal 923
1805 CTTCCAAAACGAACAGGACACACCAAAATTCGCCCTCTGCTAGT 1854
923 hrGlnLysAlaAspGlyLysGly..... 930
1855 TATGCAAAATATCCCAAGAGGAGCTCCCAATTTCTCAGGCGCAGTGTAC 1904
931 ...AlaAspValLysIleGlyAlaLys.....Th 939
1905 AGCCCTGATGATCAGTCAATGGAAGAAACAGTTACC...TTGCAACTAC 1951
939 rSerValIleLysAspHisAsnGlyLysLeuPheThrGlyLysAspLeuL 956
1952 TGGATAATGGAGCAGTGTCTGCTACTAAGGATCAGCT..... 1992

```

```

956 ysAspAlaAsnAsnGlyAlaThrValSerGluAspAspGlyLysAspThr 972
1993 .....GTCTACTCAAGGTATTTCACAACTATGACACGAATGG 2030
973 GlyThrGlyLeuValThrAlaLys.....ThrValIleAspAlaValAs 987
2031 TAGATACAGTGTAAAGTGGGCTCTGGGA..... 2061
987 nLysSerGlyTrpArgValThrGlyGluGlyAlaThrAlaGluThrGlyA 1004
2062 .....GGAGTTAAGCAGCAGCAGGAGTGTATCCCAAGCAGAGTGA 2106
1004 laThrAlaValAsnAlaGlyAsnAlaGluThrValThrSerGlyThrSer 1020
2107 GCCTGTACATACCTGGCTGGATTGAGATGATGAAATACAATGGAATCC 2156
1021 ValAsnPheLysAsnGly.....AsnAl 1028
2157 ACCAAGACCTGAAATTAATAAGCAT.....G 2182
1028 aThrThrAlaThrValSerLysAspAsnGlyAsnIleAsnValLysTyrA 1045
2183 ATGTT.....CAACACAAGCAAGTGTGT 2205
1045 spValAsnValGlyAspGlyLeuLysIleGlyAspAspLysLysIleVal 1061
2206 TTCAGCAGAACATCC.....TCGGGAGGCTCATTTGTG..... 2238
1062 AlaAspThrThrThrLeuThrValThrGlyGlyLysValSerValProAl 1078
2239 .....GCTTCTGATCTCCCAATGCTCCCATACCTGCTCTTCCACCTG 2284
1078 aGlyAlaAsnSerValAsnAsn..... 1085
2285 GCCAAATCACCGACCTGAAGCGGAAATTCACGGGCGAGTCTCATTAAT 2334
1086 ..AsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsnAsn 1101
2335 CTGACTTGGACAGCTCTCTGGGATGATTATGACCATGGAACAGCTCAAA 2384
1102 LeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyGluSerGluL 1118
2385 GTATATC..... 2391
1118 yGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysAlaG 1135
2392 .....ATTCCAATAAGTACAAAGTATTCCTGATCTCAGAGACAAGTTC 2433
1135 yLysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSerLeu 1151
2434 AATGAATCTCTCAAGTGAATACTACTGTCTCATCCCAAGGAAGCCAA 2483
1152 GlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAlaAs 1168
2484 C.....TCTGAGGAAGTCTTTTGTGTTAAACCAAGAAACATTAATCTTT 2526
1168 nGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrIle 1185
2527 .....GAAAATGGCACAGAT 2541
1185 hrLeuAlaAsnGlyAlaAlaAlaGlyThrAspAlaSerAsnGlyAsnThr 1201
2542 CTTTTCATGCTATTTCAGGCTGTTGATAA..GGTCGATCTGAATCAGAAA 2590
1202 IleSerValThrLysAspGlyIleSerAlaGlyAsnLysGluIleThras 1218
2591 TATCCAAATGTCAGCAGTATCTTTGTTTATTCCTCCACAGACTCCGCCA 2640
1218 nValLys.....SerAlaLeuLysThrTyrLysAspThrGlnAsnThrA 1233
2641 GAGACACCTAGTCTCTGATGAACGCTGCTGCTCTTGTCTTAATATTCATAT 2690
1233 laAspGlu..... 1235

```



```
335 lySerAspIleAlaLysThrGlyGlyPhe.....ValGIuThrSer 348
|||||
1738 TCACAAACCTTGACCCCTG..... 1755
|||||
349 GlyHisAspLeuSerIleGlyAspValIleValAspAlaLysGIuTr 365
1756 .....ACTGTCACCTCCGTCGCT 1774
|||||
365 pLeuLeuAspProAspAspValSerIleGIuThrLeuThrSerGlyArgA 382
1775 CCAATGCTACCTCCCTCCCAATTACAGTGACTTCCAAACGACAAAGGAC 1824
|||||
382 snAsnThrGlyGluAsnGlnGlyTyrThrGlyAspGlyThrLysGlu 398
1825 ACCAGCAAA.....TTCCCCAGCCCTCTGCTAGTTTATGCCAATAT 1865
|||||
399 SerProLysGlyAsnSerIleSerLysProThrLeuThrAsnSerThrLe 415
1866 TCGCCAAGGACCTCCCAATTCTCAGGGCCAGTGTACAGCCCTGATTG 1915
|||||
415 uGluGln.....IleLeuArgArgGlySerTyrValAsnIleT 428
1916 AATCAGTGAATGGAACACACTTACCTTGGAACTA...CTGGATAATGGA 1962
|||||
428 hrAlaAsnAsnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGly 444
1963 GCAGGTGCTGATGCTACTAAGGATGACGGTGTCTTACTCAAGGTATTTCCAC 2012
|||||
445 SerLeuThrLeuHisThrLysArgAspGlyValLysIleAsnGlyAspI 461
2013 AACTTATGACACGAATGTTAGATACAGTGTAAAGTGGCGGCTCTGGAG 2062
|||||
461 eThrSerAsnGluAsnGlyAsnLeuThrIleLysAlaGlySerTrpVala 478
2063 GAGTTAAGCAGCAGCAGCAGAGTATACCCAGCAGAGTGGAGCAGCTG 2112
|||||
478 spValHis.....LysAsnIleThrLeuGlyThrGlyPheLeu 490
2113 TACATACCTGGCTGGATTGAGATGATAAATACAAATGGAATCCACCAAG 2162
|||||
491 AsnIle.....ValAlaGlyAspSerValAlaPheGluArgGlu 504
2163 ACCTGAA..... 2169
504 yAspLysAlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrI 521
2170 .....ATTAATAAGGATGATTCAACACAAG.....CAAGTGTGTTTC 2208
|||||
521 leThrValAsnLysAspAspLysGlnPheArgPheAsnAsnValSerIle 537
2209 AGCAGAACATCCTCGGAGGCTCATTTGTGGCTTCTCATGTCCTCCCAATGC 2258
|||||
538 AsnGlyThrGlyLysGlyLeuLysPheIleAlaAsnGln..... 550
2259 TCCCATACCTGATCTCTTCCCACTGGCCAAATCAC.....GACC 2299
|||||
551 .....AsnAsnPheThrHisLysPheAspG 559
2300 TGAAGCGGAAATTCAGGGGAGCTCATTAATCTGACT..... 2340
|||||
559 lyGluIleAsnIleSerGlyIleValThrIleAsnGlnThrThrLysLys 575
2341 .....TCGACAGCTCTCGGGGATGATTATGACCAT..... 2370
|||||
576 AspValLysTyrTrpAsnAlaSerLysAspSerTyrTrpAsnValSerSe 592
2371 .....GGACAGCTCACAAATAT.....ATCATTCGAA 2398
|||||
592 rLeuThrLeuAsnThrValGlnLysPheThrPheIleLysPheValAsp 609
2399 TAAGTACAAGTATCTTGATCTCAGACAAAGTTCAATGAATCTCTCAA 2448
|||||
```

```
609 erGlySerAsnSerGlnAspLeuArgSerSerArgArgSerPheAlaGly 625
2449 GTCATACTACTGCTCTCATCCCAAGGAA.....GCCAA 2483
|||||
626 ValHisPheAsnGlyIleGlyGlyLysThrAsnPheAsnIleGlyAlaAs 642
2484 CTCTGAGGAAGTCTTTTGTTTAAACACGAGAAACATTACT 2523
|||||
642 nAlaLysAlaLeuPheLysLeuLysProAsnAlaAlaThr 655
seq_name: /SID52/gcgdata/geneseq/geneseqp/AA1993.DAT:AA41725
seq_documentation_block:
ID AA41725 standard; Protein; 1536 AA.
XX
AC AA41725;
XX
DT 26 APR-1994 (first entry)
XX
DE High molecular weight protein 1 (HMW1).
XX
KW HMW; high molecular weight protein; virus; vaccine; influenza;
KW epitope; immunity; haemophilus influenzae; gene cluster.
XX
OS Haemophilus influenzae.
XX
FH Key Location/Qualifiers
FT Misc-difference 668..677
FT /note= "Possibly incorrect sequence. Alternative
FT sequence for this region is LNVSEGEFN.
FT (See comments)."
```

High molecular weight surface proteins - of non-typeable
haemophilus which exhibit immunogenic properties

Claim 3; Figure 2/10; 100pp; English.

The isolation and purification of the high molecular weight protein enables the identification of the major protective epitopes of the protein by conventional epitope mapping. These epitopes can then be synthesised using standard techniques and incorporated into fully synthetic or recombinant vaccines. This sequence is claimed to be the same as that given in AA41723 (high molecular weight protein 1) although it does differ slightly. (Repeated regions which are possibly incorrect and occur in the corresponding nucleotide coding sequence contribute to these differences).

Sequence 1536 AA;

alignment_scores:
Quality: 139.50 Length: 938
Ratio: 0.319 Gaps: 49
Percent Similarity: 46.588 Percent Identity: 20.256

alignment_block:
-US-091949-696-18 x AA41725 ..

Align seg 1/1 to: AAR41725 from: 1 to: 1536

```
217 TTGAGCTACA.....GGAAGCGATTTTATTTCAAAAATGT 254
|||||:|||||
631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnIle 643
|||||:|||||
255 TGCCATTTTGTATCTCTGAACATGAAGCAAGAGCTGCTATGTGAGAC 304
|||||:|||||
643 eSerMetValLeuPro.....LysAsnGluSerGlyTyrAspLysp 657
|||||:|||||
305 CAAAACTTGAGACCTACAAAAT..... 327
|||||:|||||
657 heLysGlyArgThrTyrTrpAsnLeuThrSerLysValAspMetIleAsn 673
|||||:|||||
328 ...GCTGATGTTCTGGTGTGCTGAGTCTACTCTCCAGCTAATGAT..... 369
|||||:|||||
674 SerLysAspAlaLeuThrIleAspSerArg.....GlySerAspSerAl 688
|||||:|||||
370 .....GNACCTACACT.....G 382
|||||:|||||
688 aGlyThrLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysA 705
|||||:|||||
383 AGCAGATGGCAACTGTGGAGAGA.....AGGTGAAGGATCCACCTC 426
|||||:|||||
705 spThrThrPheAsn.ValGluArgAsnAlaArgValAsnPheAspIleLy 721
|||||:|||||
427 ACTCTGTATTCATTTGAGGAAAAAGTTAGCTGAATATGACCACAAAG 476
|||||:|||||
721 sAlaProIleGlyIleAsnLysTyrSerSerLeuAsnTyrAlaSerPheA 738
|||||:|||||
477 TAGGCATTTGCTCCATGATGGGCTCATCTACGATGGGAGATTTGACG 526
|||||:|||||
738 sNGLysIleSerValSerGly.....GlyGlySerValAsp 750
|||||:|||||
527 AGTACAATAATGATGAGAAATCTACTTATCCAAATGGAAGATACA.... 572
|||||:|||||
751 PheThrLeuLeuAlaSerSerSerAsn.ValGlnThrProGlyValValI 767
|||||:|||||
573 ..AGCAGTAAGATGTTACGAGGATTTACTGTACAAATGTAGTAAGAA 620
|||||:|||||
767 leAsnSerLysTyrPheAsn.ValSerThrGlySerSerLeuArgPheLy 783
|||||:|||||
621 GTGTCAGGA.....GCAGCTGTACACCAAAAGATCCACAT 658
|||||:|||||
783 sThrSerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrL 800
|||||:|||||
659 TCAAT.....AAAGTAACAGGACTCTATGAAAAAGGATGTGAG 696
|||||:|||||
800 euAsnAlaThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAsp 816
|||||:|||||
697 TTTGTTTCCAAATCCCGCCAGCAGGAGAGGCTCTCTATAATGTTT.... 741
|||||:|||||
817 GlyMetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluG 833
|||||:|||||
742 .....GCACAACATGTTGATTTCTA 760
|||||:|||||
833 yGlyLysMetArgPheGlySerArgLysAlaValThrGluIleGluGlyA 850
|||||:|||||
761 TAGTTGAATCTGTACAGAA..... 780
|||||:|||||
850 snValThrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPhe 866
|||||:|||||
781 CAAACACACACAAGAAGCTCCAAACAAAGCAAAAT.....CAAAA 821
|||||:|||||
867 AspAsnHisGlnLysProLeuThrIleLysLysAspValIleIleAsnSe 883
|||||:|||||
822 ATGCAATCTCGAAGCACATGGGAAGTG..... 849
|||||:|||||
883 rGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuT 900
|||||:|||||
850 ..ATCCGTGATTTGAGGACTTTTAAAGAAACCACTCTCTATGACA.... 891
|||||:|||||
```

```
960 hrValGluSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsn 916
|||||:|||||
892 .....ACACAGCCACCAAAATCCACCTTCTCTATTCGTGCTCA 926
|||||:|||||
917 ValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLy 933
|||||:|||||
927 GATTGGACAAAGAATTGTGTGTTTAGTCTTGTACAAATCTGGAAGCATGG 976
|||||:|||||
933 sGlyIleAlaArgPheLysAspIleAspAsnSerLysAsnLeuSerIleT 950
|||||:|||||
977 CGACTGGTAAACCGCTC.....AATCGACTG 1002
|||||:|||||
950 hrThrAsnSerSerThrTyrArgThrIleIleSerGlyAsnIleThr 966
|||||:|||||
1003 AATCAAGCAGCCAGCTTTTCCTGCTGCAGACAGATTGAGCTGGGCTCTG 1052
|||||:|||||
967 AsnLysAsnGlyAspLeuAsnIleThrAsnGlu.....GlySer.. 979
|||||:|||||
1053 GTTGGGATGTTGACATTTGACAGCTGCTGCCATGTACAAAGTGAACCTCA 1102
|||||:|||||
980 .....AspThrGluMetGlnIleGlyGlyAspValS 990
|||||:|||||
1103 TACAG.....ATAAACAGTGCAGCTGACAGGACACACTCCCC 1140
|||||:|||||
990 erGlnLysGluGlyAsnLeuThrIleSerSerAspLysIleAsnIleThr 1006
|||||:|||||
1141 AAAAGA.....TTACCTGCAGCAGCTTACAGAGGAGCTCCATCTGCAG 1184
|||||:|||||
1007 LysGlnIleThrIleLysAlaGlyValAspGlyGluAsnSerSerAs 1023
|||||:|||||
1185 CGGCTTCGATCGGCATTTACTGTGATTAGGAAGAA..... 1221
|||||:|||||
1023 pAlaThrAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuT 1040
|||||:|||||
1222 .....TATCCAACT 1230
|||||:|||||
1040 hrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLys 1056
|||||:|||||
1231 GATGGATCTGAAATTTGCTGCTGCTGACGGATGGGAAGACACACTATAAG 1280
|||||:|||||
1057 AspGlySerAsp.....LeuThrIleGlyAsnThrAsnSerAlaAs 1070
|||||:|||||
1281 TGGG.....TGCTTTAAGGAGTCAACAAAGT.... 1308
|||||:|||||
1070 pGlyThrAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysI 1087
|||||:|||||
1309 ..GGTGGCCATCATCCACACAGCTCGCTTTTGGGGCCCTCTGCAGCTCAAGAA 1356
|||||:|||||
1087 leSerAlaAspGlyHisLysValThrLeuHis..... 1097
|||||:|||||
1357 CTAGAGGAGCTGTCCAAAATG....ACAGGAGTTTACAGACATATGCTTC 1403
|||||:|||||
1098 .....SerLysValGluThrSerGlySerAsnAsnThrGln 1110
|||||:|||||
1404 AGATCAAGATTCAGAACAAATGGCTCATTGATGCTTTTGGGCCCTTTTCAT 1453
|||||:|||||
1110 uAspSerSerAspAsnAsnAla..... 1117
|||||:|||||
1454 CAGGAAATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTTGAGAGTAAGGA 1503
|||||:|||||
1118 .....GlyLeuThrIleAspAlaLysAsn 1125
|||||:|||||
1504 TTAACCTCCAGAACAGCCAGTGGATGAATGGCAGAGTATGCTGGACAG 1553
|||||:|||||
1126 ValThrValAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAl 1142
|||||:|||||
1554 CACCGTGGAAAGGACACTTTGTTTCTTATCACCTGGACAACGCGCTC 1603
|||||:|||||
1142 aThrSerGlyGlu.....IleThrThrLysThr..... 1151
|||||:|||||
1604 CCCAAATCTTCTCTCGGATCCCATGGGACAGCAAGGAGTGGCTTGTGA 1653
|||||:|||||
1152 .....GlyThrThr 1154
|||||:|||||
```



```
1654 GTGGACAAAACACCAAAATGGCCCTACCTCCAAATCCAGCGCATTCGTAA 1703
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  |||
1155 ileasnAlaThrThrGly.....AsnValGluIleThrAlaG1 1167
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  |||
1704 GGTGGCAGCTTGGAAATACAGTCTGCAAGCAAGCTCACAACCTTGACCC 1753
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  |||
1167 nThrGlySerIleLeuGlyGlyIleGlySerSerSerGlySerValThrL 1184
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  |||
1754 TGACTGTCAGTCCCGGCG.....TCCAATGCTACCCCTGCCCTCCA 1794
      |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||
1184 eutThrAlaThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThr 1200
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  |||
1795 ATTACAGTGACTTCCAAAACGAACAGGACACACAGC..... 1830
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  |||
1201 ValThrValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerTh 1217
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  |||
1830 ..... 1830
1217 rIleLysGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleG 1234
1831 ..... 1831
1234 lyGlyThrIleSerGlyThrValGluValLysAlaThrGluSerLeu 1250
      |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||
1849 GTAGTTTATCAATATTCG.....CAAGGACCTCCCAAT 1886
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  |||
1251 ThrThrGlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnVa 1267
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  |||
1887 TCTCAGGCGCAGTGTCAACCCCTGATTAATGATCAGTGAATGGAAACAG 1936
      |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||
1267 lThrSerAlaThrGlyThr...lleGlyGlyThrIleSerGlyAsnThrV 1283
1937 TTACCTTGGAA.....CTACTGGATATAGGACAGGT 1968
      |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||
1283 alaAsnValThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGlu 1299
1969 GCTGATGCTACTAAG.....GATCAGCGTGTCTA 1997
1300 ileasnAlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLe 1316
1998 CTCAAGGTATTTCACAACTTATCACAGCAATGGTAGATACAGTGTAAAG 2047
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  |||
1316 uThrThrGluAlaSerSerHisIleThrSerAlaLysGlyGlnValAsnL 1333
2048 TGGGGGCT.....CTGGGAGGAGTTACGCGACGACGAGCG 2082
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  |||
1333 euSerAlaGlnAspGlySerValAlaGlySerIleAsnAlaAla..... 1347
2083 AGAGTGATACCCAGCAGAGTGGAGCACTGTACATACCTGGCTGGATTGA 2132
1348 AsnValThrLeuAsnThrThrGlyThrLeuThrThr.....Vally 1361
2133 GAATGATGAATAACAATGGAATCCACCAGACCTGAAATTAATAAGGATG 2182
      |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||
1361 sGlySerAsnIleasnAlaThrSerGlyThrLeuValIleAsnAlaLysA 1378
2183 ATGTTCAA.....CACAGCAAGTGTGTTTC 2208
1378 spAlaGluLeuAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAla 1394
2209 AGCAAGACATCTCGGAGGCTCATTTGTGCTTCTCATGTCCTCCCAATGC 2258
1395 ThrAsnAlaAsnGlySerGlySerValIleAlaThrThrSerSerArgVa 1411
2259 TCCCATACCT...GATCTCTTCCACCTGGCCAAATC..... 2292
1411 lAsnIleThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleIleSerL 1428
2293 ..ACCGACCTGAAGCGGAAATTCACGGGGGCAGTCTCATTAATCTGACT 2340
1428 ysAsnGlyIleAsnThrValLeuLeuLysGlyValLysIleAspValLys 1444
```

```
2341 TGCACAGCTCTCGG...GATGATTATGACCATGGAAACAGCTCACAGTA 2387
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  |||
1445 TyrIleGlnProGlyIleAlaSerValAspGluValIleGluAlaLys.. 1460
2386 TATCATTCGAATAAGTACAAGTATTTCTTGATCTCAGAGACAAGTTCAATG 2437
      |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||
1461 .....ArgIleLeuGluLysValLysAspLeuSerAspGluGluArg 1475
2438 AATCTCTT 2445
      |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||:::  |||
1475 luAlaLeu 1477
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW56322

seq_documentation_block:

ID AAW56322 standard; Protein; 2039 AA.

XX

AC AAW56322;

XX

DT 19-AUG-1998 (first entry)

XX

DE Haemophilus paragallinarum antigenic protein #2.

XX

KW Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;

KW vaccine; chicken infectious coryza; CIC; fowl.

XX

OS Haemophilus paragallinarum.

XX

FH Key Location/Qualifiers

FT Peptide 1..70

FT /label= signal

FT Protein 71..2039

FT /note= "antigenic protein"

XX

PN WC3812331-Al.

XX

PD 26-MAR-1998.

XX

PF 12-SEP-1997; 97WO-JP03222.

XX

PR 19-SEP-1996; 96JP-0271408.

XX

PA (KACA) 2H KAGAKU & KESSEI RYOHO KENKYUSHO.

XX

PA (KACA) CEMO-SERO-THERAPEUTIC RES INST.

XX

PI Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;

XX

DR WPI; 1998-230318/20.

XX

DR N-PSDB; AAV22837.

XX

FS Claim 5; page 71-87; 108pp; Japanese.

XX

CC The present sequence represents an antigenic protein derived from

CC Haemophilus paragallinarum strain C-53-47. The antigenic protein

CC stimulates the production of HI antibodies in fowl. The protein

CC and DNA coding for it can be used in the preparation of vaccines

CC for the prevention of chicken infectious coryza (CIC). The protein

CC and its antibodies can be used in the diagnosis and treatment of CIC.

XX

SQ Sequence 2039 AA;

alignment_scores:

Quality: 139.50

Ratio: 0.388

Percent similarity: 45.512

Length: 791

Gaps: 40

Percent identity: 19.975

alignment_block:

US-09-049-696-18 x AAW56322

Align seg_1/1 to: AAW56322 from: 1 to: 2039

871	AAGAAAACCACTCTCTATGACAACACAGCCACCAATCCC.....	909
1179	LysAlaAsnSerProIleThrValGluProSerThrAspAsnAsnLysLys	1195
910ACWTCATCTGCTGCAGATTGGCAAAACAATTGTCGTT	949
1195	sLysThrPheThrValGlyLeuMetLysAspIleGlyValAsnSerI	1212
950	TAGTCCTTGCAAAATCTGGA.....AGCATGGCACTGGTAACCCG	990
1212	leThrPheAspLysSerGlyGlnAspLeuAsnGlnValThrGly.....	1226
991	CTCAATCGACTGAATCAAGCAGGCCAGCTTTCTCTGTCGACAGATTGA	1040
1227ArgMetSerSerAlaGlyLeuThrPheLysLysGlyAspThr	1241
1041	GCTGGGCTCTGGGTGGGATG.....CTGACATTGACAGTG	1078
1241	rAsnGlySerThrThrPheAlaGluAspGlyLeuThrIleAspSerI	1258
1079	CTGCCCCAT...GTACAAAGTGAATCATACAGATAAAC.....	1113
1258	hrThrAsnSerAlaGlnThrAsnLeuValLysValSerArgAspGlyPhe	1274
1114AGTGCAGTGCAGGACACACTC.....	1137
1275	SerValLysAsnGlySerAspLysSerLysLeuAlaSerThrLysLeuSe	1291
1138GCCAAAAGATTACCTGCAGCAGCTTCAGGA.....	1167
1291	rIleGlyAlaGluAsnAlaGluHisValGluValThrLysSerGlyIleA	1308
1167	1167
1308	laLeuLysAlaAspAsnThrSerAspLysSerSerIleThrLeuAlaGln	1324
1168GGAGCTCCATC...TGCAG	1184
1325	AspAlaIleThrLeuAlaGlyAsnAlaThrGlyThrAlaIleLysLeuTh	1341
1185	CGGCTCGCATCGGCATTACTGCTATTAGGAAGAAATATCCAACATGATG	1234
1341	rGlyValAlaAspGlyAsnIleThrValAsnSerLysAspAlaValAsnG	1358
1235	GATCTGAAATT.....GTGCTGCTCAGCGATGGG.....	1263
1358	lyGlyGlnLeuArgThrLeuLeuGlyValAspSerGlyAlaLysIleGly	1374
1264GAAGACACACTATAGTGGTGCTTTAAGCAGGTCAACAAAG	1307
1375	GlyThrGluLysThrThrIleSerGluAlaIleSerAspValLysGlnAl	1391
1308	TGGTGCCATCATCCACAGTCGCTTTGGGGCCCTCTGCAGTCAAGAAC	1357
1391	a.....LeuThrAspAlaThrLeuAlaThrLysAlaAspAsnLysA	1405
1358	TAGAGGACCTGCCAAATGACAGAGGTTTACAG.....ACA	1395
1405	snGlyLysThrValLysLeuThrAspGlyLeuAsnPheThrSerThrThr	1421
1396	TATGCTTCAGATCAAGTTCAGAACATGGCCCTCATTTGAT.....	1434
1422	AsnIleAspAlaSerValGluAspAsnGlyValValLysPheThrLeuL	1438
1435GCTTTTGGGGCCCTTTCATCAG	1456
1438	sAspLysLeuThrGlyLeuLysThrIleAlaThrGluSerLeuAsnAlaA	1455
1457	GAATGGAGCTGCTCTCAGCGCTCATCCAGCTT.....GAGAGTAAG	1500

[illegible]

```
2032 AGATACAGTGTAAAGTGGGGCTCTGGAGGAGCTTAACGCAGCAGAGC 2081
   ::::: ||| ||::: ||
1765 GLuLysAsnMetLeuVal.....LeuSerAspLysLysAlaLysProAr 1779
   ::::: ||| ||::: ||
2082 GAGAGTGATACCCAGCAGAGTGGAGCTGTACATACCTGGCTGGATTG 2131
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
1779 gAlaValLeuAspGLyGlnAsnGlyAlaLeuThrLeuValGly..... 1793
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
2132 AGAATGATGAATAACAA.....TGAATCCCAAGACCTGAAATT 2172
   ||::: ||| ||::: ||::: ||::: ||::: ||::: ||
1794 ..AsnAspAspSerGlnValThrLeuSerSerLysLysGlyLysAspIle 1809
   ||::: ||| ||::: ||::: ||::: ||::: ||::: ||
2173 ATATAGGATGTTCAACACAGCAAGTGTCTTCACGACAGAACATCTCTC 2222
   ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||
1810 AspGLyAsnAspLeuSerArgLeuSerValThrThrGluArgThrAsnAl 1826
   ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||
2223 GGGAGGCTCATTTGTGGCTCTGTGATGTCCTCCAAATGCTCCCATCTGATC 2272
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
1826 aAspGLyGlnLeuGluLysValGluThrSerPheAlaThrMetAspAsp. 1842
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
2273 TCTTCCCACTGGCCAAATCACCGACTGAAGCGGGAATTCACGGGGC 2322
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
1843 .....GlyLeuLysPheLysAlaAspGLyAsp 1851
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
2323 ACTCTCATTAATCTG.....ACTTGGACAGCTCTCTGGGATGA 2360
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
1852 LysValIleAsnLysLysLeuAsnGluThrValGluIleValGlyAspGl 1868
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
2361 TTATGACCATGCAACAGCTCACAAAGTATATCATTCGAATAACTACAAGTA 2410
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
1868 u.....AsnValThrThrSerI 1874
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
2411 TTCTTGAT.....CTCAGAGACAAGTTCATGAATCTCTCAAGTG 2451
   || ||| ||::: ||::: ||::: ||::: ||::: ||::: ||
1874 leThrAspAspAsnLysValLysValSerLeuAsnLysLysIleAlaIle 1890
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
2452 ATACTACTGCTCTCATCCCAAGGAAGCAACTCTGAGGAGCTCTTTT 2501
   ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||
1891 AspGluValLysIleProAsnThrAspPronAspAlaGln..... 1903
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
2502 GTTTAAACAGAAAACATTACTTTTGAAATGGCACAGATCTTTTCATTG 2551
   ||| ||::: ||::: ||::: ||::: ||::: ||::: ||
1904 ....LysGLyAspSerIleValIleAsnAsnGly.....G 1914
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
2552 CTATTCAGGCTGTTGATAGGTC 2574
   ::::: ||::: ||::: ||::: ||::: ||::: ||::: ||
1914 lYleHisAlaGlyAsnLysVal 1921
   ||::: ||::: ||::: ||::: ||::: ||::: ||::: ||
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT: AAB01847

seq_documentation_block:

ID AAB01847 standard; Protein: 1095 AA.

AC AAB01847;

DT 11-SEP-2000 (first entry)

DE Haemophilus influenzae strain 12 mature HMW1A protein, SEQ ID NO:69.

KW Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.

OS Haemophilus influenzae strain 12.

PN WO200020609-A2.

PD 13-APR-2000.

XX 07-OCT-1999; 99WO-CA00938.

PF

```
XX 07-OCT-1998; 98US-0167568.  
PR 08-DEC-1998; 98US-0206942.  
XX  
XX  
XX (CONN-) CONNAUGHT LAB LTD.  
XX  
XX Loomore SM, Yang Y, Klein MH;  
PI WPI; 2000-303789/26.  
DR N-5SDB; AAA52196.  
XX  
XX Nucleic acid molecule for producing recombinant high molecular weight  
PT proteins of Haemophilus which are used as a vaccine to provide  
PT protection against Haemophilus induced diseases in humans -  
XX  
XX Claim 8; Fig.28F-Q; 307pp; English.  
XX
```

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwABC operon from a non-typeable (non-encapsulated) H. influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene clusters termed hmwLABC and hmw2ABC. Each hmwABC operon comprises hmwA, hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins and the hmwB and hmwC genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMW proteins. The modified hmwABC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMW. The invention also discloses hmwA genes (AAA52175-A52198) and HMW proteins (AAB01824-B01849) from the non-typeable H. influenzae strains Joyce, KI, K21, LCD2, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant H. influenzae HMW proteins which can be used as vaccines to mediate a humoral or cell-mediated immune response to provide protection against diseases in humans caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HMW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus, HMW proteins and/or HMW peptides. The nucleotide sequences encoding the HMW proteins can be used to isolate and clone hmw genes from other non-typeable strains of Haemophilus via hybridisation reactions. The present sequence represents a mature HMW protein from a non-typeable strain of H. influenzae.

XX Sequence 1095 AA;

alignment_scores:
Quality: 138.50 Length: 936
Ratio: 0.318 Gaps: 49
Percent Similarity: 46.581 Percent Identity: 20.406
alignment_block:
US-09-049-696-18 x AAB01847 ..
Align seg 1/1 to: AAB01847 from: 1 to: 1095
217 TTTGAAGCTACA.....GGAAGCGATTTTATTTCAAAAATGCT 254
|||||:|||||
190 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnIle 202
255 TGCCATTTCATTCCTGAACATGGAAGCGTGAATGTGAGAG 304
|||||:|||||
202 eSerMetValLeuPro.....LysAsnGluSerGlyTyrAspLysP 216
305 CAAAATTCAGACCTACAAAATGCTGATGTTCTG...GTTCTGAGTCT 351
|||||:|||||
196 heLysGlyArgThrTyrTrpAsnLeuThrSerLeuAsnValSerGluSer 232
352ACTCTCCAGGTAATGAT..... 369
196 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspAlaGlyTh 249


```

778 sGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyClyT 795
1831 .....AAATCCCCAGCCCTCTGGTAGTT 1854
795 hrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThrTh 811
1855 TATGCCAAATATTCG.....CAAGGAGCCTCCCAATCTCAG 1892
812 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 828
1893 GGCCAGTGTCCAGCCCTGATTGAATCAGTGAATGGAACACAGTTACCT 1942
828 rAlaThrGlyThr...IleGlyThrIleSerGlyAsnThrValAsnV 844
1943 TGGAA.....CTACTGGATAATGAGCAGGTGCTGAT 1974
844 alThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 860
1975 GCTACTAAG.....GATGACGGTGTCTACTCAAG 2003
861 AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 877
2004 GTATTTTCACAACTTATCACACGAATGTAGATACAGTGTAAAGTGGGG 2053
877 rGluAlaSerSerHisIleThrSerAlaLysGlyGlnValAsnLeuSerA 894
2054 CT.....CTGGGAGGAGTTAACCCAGCCAGCAGGAGAGTG 2088
894 lagInAspGlySerValAlaGlySerIleAsnAlaAla.....AsnVal 908
2089 ATACCCAGCAGAGTGGAGCAGTGTACATACCTGGTGGATTGAGAATGA 2138
909 ThrLeuAsnThrThrGlyThrLeuThrTh.....ValLysGlySe 922
2139 TGAATACATCAATGAATCCACCAAGACCTGAAATTAATAGGATGATGTT 2188
922 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 939
2189 AA.....CACAGCAAGTGTGTTTCAGCAGA 2214
939 luLeuAsnGlyAlaAlaLeuLysGlyAsnHisThrValValAsnAlaThrAsn 955
2215 ACATCTCGGAGGCTCATTTGGCTCTGATGTCCTCCCAATGCTCCCAT 2264
956 AlaAsnGlySerGlySerValIleAlaThrThrSerSerArgValAsnII 972
2265 ACCT...GATCTCTTCCACCTGGCCAAATC.....ACCG 2296
972 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleIleSerLysAsnG 989
2297 ACCTGAAGCGGGAATTCAGGGGAGTCTCATTAATCTGATCTGGACA 2346
989 lyIleAsnThrValLeuLeuLysGlyValLysIleAspValLysTyrIle 1005
2347 GCTCTCTGG...GATGATTATGACCTGGAACAGCTCACAGTATATCAT 2393
1006 GlnProGlyIleAlaSerValAspGluValIleGluAlaLys..... 1019
2394 TCGAATAAGTACAGTATCTTGATCTCAGAGCAAGTTCATTAATGATCTC 2443
1020 .ArgIleLeuGluLysValLysAspLeuSerAspGluGluArgGluAlaL 1036
2444 TT 2445
1036 eu 1036

```

seq_name: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT:AA41723

seq_documentation_block:

ID AA41723 standard; Protein; 1536 AA.

XX

AC

XX

AA41723;

XX

```

26-APR-1994 (first entry)
XX
DE High molecular weight protein 1 (HMW1).
XX
KW HMW; high molecular weight protein; virus; vaccine; influenza;
KW epitope; immunity; haemophilus influenzae.
XX
OS Haemophilus influenzae.
XX
PN WO319090-A.
XX
PD 30-SEP-1993.
XX
PF 16-MAR-1993; 93WO-US02166.
XX
PR 16-MAR-1992; 92GB-0005704.
XX
PA (BARE/) BARENKAMP S J.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Barenkamp SJ.
XX
DR WP1; 1993-320683/40.
XX
DR N-PSDB; AAQ49506.
XX
PT High molecular weight surface proteins - of non-typeable
PT haemophilus which exhibit immunogenic properties
XX
PS Claim 3; Figure 2; 100pp; English.
XX
CC The isolation and purification of the high molecular weight protein
CC enables the identification of the major protective epitopes of the
CC protein by conventional epitope mapping. These epitopes can then be
CC synthesised using standard techniques and incorporated into fully
CC synthetic or recombinant vaccines.
XX
SQ Sequence 1536 AA;

```

alignment_scores:

Quality: 138.50 Length: 936
 Ratio: 0.318 Gaps: 49
 Percent Similarity: 46.581 Percent Identity: 20.406

alignment_block:

```

US-09-049-696-18 x AA41723 ..
Align seg 1/1 to: AA41723 from: 1 to: 1536
217 TTTCAGCCACAC.....GGAAACCGATTATTTCACAAATGT 254
|||||:|||||
631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnII 643
255 TGCCATTTGATCTCTCAACATGGAAGACAAAGCGCTGACTATGTGAGAC 304
|||||:|||||
643 eSerMetValLeuPro.....LysAsnGluSerGlyTyrAspLysP 657
305 CAAACTTGAGACCTACAAATGCTGATGTTCTG...GTTGCTGAGTCT 351
|||||:|||||
657 heLysGlyArgThrTyrTTPAsnLeuThrSerLeuAsnValSerGluSer 673
352 .....ACTCTCCAGGTAATGAT..... 369
|||||:|||||
674 GlyGluPheAsnLeuThrIleAspSerArgGlySerAspSerAlaGlyTh 690
370 .....GAACCTTACACT.....GAGCAGA 388
|||||:|||||
690 rLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThrT 707
389 TGGGCAACTGTGGAGAGA.....AGGGTGAAGGATCCACCTCCTCT 432
|||||:|||||
707 hrPheAsn.ValGluArgAsnAlaArgValAsnPheAspIleLysAlaPr 723

```

433 GATTTCATTCGAGGAAAGAGTTAGCTGAATATGACACCAAGTAGGC 482
 |||:|||||:||||| ||| ||||| |||:
 723 oileglyIleAsnLysTyrSerSerLeuAsnTyrAlaSerPheAsnGlyA 740
 |||:|||||:||||| ||| ||||| |||:
 483 ATTTCATTCAGTGGGCTCATCGATGAGGAGTATTTACAGAGTACA 532
 |||:|||||:||||| ||| ||||| |||:
 740 snIleSerValSerGly.....GlyGlySerValAspPheThr 752
 |||:|||||:||||| ||| ||||| |||:
 533 ATAATGATGAGAAATTCCTACTTATCCCAATGAAGAATACA.....AGCA 576
 |||:|||||:||||| ||| ||||| |||:
 753 LeuLeuAlaSerSerSerAsn.ValGlnThrProGlyValValIleAsnS 769
 |||:|||||:||||| ||| ||||| |||:
 577 GTAAGATTCAGCAGGATTTACTGTCGACAAATAGTAGTAAGAAGTCA 626
 |||:|||||:||||| ||| ||||| |||:
 769 erLysTyrPheAsn.ValSerThrGlySerSerLeuArgPheLysThrSe 785
 |||:|||||:||||| ||| ||||| |||:
 627 GGA.....GCACGCTGTACACCAAAAGATGCACATTCAT. 663
 |||:|||||:||||| ||| ||||| |||:
 785 rGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeuAsnA 802
 |||:|||||:||||| ||| ||||| |||:
 664AAAGTAACAGGACTCTATGAAAAAGGATGTGAGTTGTT 702
 |||:|||||:||||| ||| ||||| |||:
 802 laThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspGlyMet 818
 |||:|||||:||||| ||| ||||| |||:
 703 CTCCAATCCCCCAGCAGGAGAGCTCTCTATATGTTT..... 741
 |||:|||||:||||| ||| ||||| |||:
 819 IleGlyLysGlyIleValAlaLysAsnIleThrPheGluGlyGlyAs 835
 |||:|||||:||||| ||| ||||| |||:
 742GCAACAACATGTTGATTCTATATGTTG 766
 |||:|||||:||||| ||| ||||| |||:
 835 nIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsnValT 852
 |||:|||||:||||| ||| ||||| |||:
 767 AATTCGTACAGAA.....CAAAC 786
 |||:|||||:||||| ||| ||||| |||:
 852 hrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPheAspAsn 868
 |||:|||||:||||| ||| ||||| |||:
 787 CACAACAAGAGCTCCAAACAGCAAAAT.....CAAATGCA 827
 |||:|||||:||||| ||| ||||| |||:
 869 HisGlnLysProLeuThrIleLysLysAspValIleAsnSerGlyAs 885
 |||:|||||:||||| ||| ||||| |||:
 828 TCTCCGAAGACATGGGAAGTG.....ATCC 853
 |||:|||||:||||| ||| ||||| |||:
 885 nLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuThrValG 902
 |||:|||||:||||| ||| ||||| |||:
 854 GTGATTCAGGACTTTAAGAAACCATCTCTATGACA..... 891
 |||:|||||:||||| ||| ||||| |||:
 902 luSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsnValGly 918
 |||:|||||:||||| ||| ||||| |||:
 892ACACAGCCACCAATCCACCTCTCATTGCTGCAGATTGG 932
 |||:|||||:||||| ||| ||||| |||:
 919 GlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLysGlyG 935
 |||:|||||:||||| ||| ||||| |||:
 933 ACAAGAATGTTGTTAGTCTTGCATAAATCTGGAAGCATGGCGACTG 982
 |||:|||||:||||| ||| ||||| |||:
 935 yAlaArgPheLysAspIleAspAsnSerLysAsnLeuSerIleThrThrA 952
 |||:|||||:||||| ||| ||||| |||:
 983 GTAACCGGCTC.....AATCGACTGAATCAA 1008
 |||:|||||:||||| ||| ||||| |||:
 952 snSerSerSerThrTyrArgThrIleIleSerGlyAsnIleThrAsnLys 968
 |||:|||||:||||| ||| ||||| |||:
 1009 GCAGGCCAGCTTCTCTGCTGCAGACAGTGTGAGTGGGCTCTGGGTGG 1058
 |||:|||||:||||| ||| ||||| |||:
 969 AsnGlyAspLeuAsnIleThrAsnGlu.....GlySer..... 979
 |||:|||||:||||| ||| ||||| |||:
 1059 GATGTGACATTTGACAGTGTGCCCATGTACAAAGTGAATCATACAG. 1107
 |||:|||||:||||| ||| ||||| |||:
 980AspThrGluMetGlnIleGlyGlyAspValSerGlnL 992
 |||:|||||:||||| ||| ||||| |||:
 1108ATAACAGTGCAGTCACAGGCACACACTCCCAAAACA 1146
 |||:|||||:||||| ||| ||||| |||:
 992 ysGluGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrLysGln 1008
 |||:|||||:||||| ||| ||||| |||:
 1147TTACCTGCAGAGCTTCAGGAGGAGTCTCCTCTGACGGGCT 1190

1009 IleThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAlaIath 1025
 |||:|||||:||||| ||| ||||| |||:
 1191 TCGATCGGCATTTACTGTGTAGTAGGAAGAAA..... 1221
 |||:|||||:||||| ||| ||||| |||:
 1025 rAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnA 1042
 |||:|||||:||||| ||| ||||| |||:
 1222TATCCAACGTGATGA 1236
 |||:|||||:||||| ||| ||||| |||:
 1042 spLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGly 1058
 |||:|||||:||||| ||| ||||| |||:
 1237 TCTGAATTTGCTGCTGACGATGGGAGACACACACTATAAGTGG... 1284
 |||:|||||:||||| ||| ||||| |||:
 1059 SerAsp.....LeuThrIleGlyAsnThrAsnSerAlaAspGlyTh 1072
 |||:|||||:||||| ||| ||||| |||:
 1285TGCTTTAAACGAGGTCAAAAGT.....GCTG 1312
 |||:|||||:||||| ||| ||||| |||:
 1072 rAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089
 |||:|||||:||||| ||| ||||| |||:
 1313 CCATCATCCACACAGTCGCTTTGGGCCCTCTGCAGCTCAAGAAGTAGAG 1362
 |||:|||||:||||| ||| ||||| |||:
 1089 laAspGlyHisLysValThrLeuHis..... 1097
 |||:|||||:||||| ||| ||||| |||:
 1363 GAGCTGTCCAAAATG...ACAGGAGGTTTACAGACATATGCTTCAGATCA 1409
 |||:|||||:||||| ||| ||||| |||:
 1098SerLysValGluThrSerGlySerAsnAsnAsnThrGluAspSe 1112
 |||:|||||:||||| ||| ||||| |||:
 1410 AGTTTCAGAACAAATGGCTCATATGCTTTTGGGCCCTTTTCATCAGGAA 1459
 |||:|||||:||||| ||| ||||| |||:
 1112 rSerAspAsnAsnAla..... 1117
 |||:|||||:||||| ||| ||||| |||:
 1460 ATGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAAGGATTAAAC 1509
 |||:|||||:||||| ||| ||||| |||:
 1118GlyLeuThrIleAspAlaLysAsnValThr 1127
 |||:|||||:||||| ||| ||||| |||:
 1510 CTCAGAACAGCCAGTGTGATGATGGACAGTGTGATGATGGACAGCACCGT 1559
 |||:|||||:||||| ||| ||||| |||:
 1128 ValAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144
 |||:|||||:||||| ||| ||||| |||:
 1560 GGAAGAGACACTTTGTTTCTTATCATCTGGACAAACGACGCTCCCAAA 1609
 |||:|||||:||||| ||| ||||| |||:
 1144 rGlyGlu.....IleThrLysThr..... 1151
 |||:|||||:||||| ||| ||||| |||:
 1610 TCCCTCTCTGGGATCCAGTGGACAGCAAGGTGGCTTTGTAGTGGAC 1659
 |||:|||||:||||| ||| ||||| |||:
 1152GlyThrThrIleAsn 1156
 |||:|||||:||||| ||| ||||| |||:
 1660 AAAAACACCAAAATGGCTACCTCCAAATCCAGGCATTTGTAAGGTGG 1709
 |||:|||||:||||| ||| ||||| |||:
 1157 AlaThrThrGly.....AsnValGluIleThrAlaGlnThrG 1169
 |||:|||||:||||| ||| ||||| |||:
 1710 CACTTGGAAATACAGTCTGCAAGCAAGCTCACAACCTTGACCTGACTG 1759
 |||:|||||:||||| ||| ||||| |||:
 1169 ysIleLeuGlyGlyIleGluSerSerGlySerValThrLeuThrA 1186
 |||:|||||:||||| ||| ||||| |||:
 1760 TCAGTCCCGTGG.....TCCATGTACCTGCTCCCTCAATTACA 1800
 |||:|||||:||||| ||| ||||| |||:
 1186 laThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
 |||:|||||:||||| ||| ||||| |||:
 1801 GTGACTTCCAAAGACACAGGACACCCAGC..... 1830
 |||:|||||:||||| ||| ||||| |||:
 1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLy 1219
 |||:|||||:||||| ||| ||||| |||:
 1830 1830
 |||:|||||:||||| ||| ||||| |||:
 1213 sglyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyTh 1236
 |||:|||||:||||| ||| ||||| |||:
 1831AAATCCCGCAGCCCTCTGTAGTT 1854
 |||:|||||:||||| ||| ||||| |||:
 1236 hrIleSerGlyThrValGluValLysAlaThrGluSerLeuThrThr 1252
 |||:|||||:||||| ||| ||||| |||:
 1855 TATGCAATATTCG.....CAAGGAGCCTCCCAATTCCTCAG 1892
 |||:|||||:||||| ||| ||||| |||:

[illegible]

```
seq_documentation_block:
ID   AAR63505 standard; Protein; 1536 AA.
```

XX	
KW	high molecular weight protein; HW1; protective vaccine; otitis;
KW	sinusitis; bronchitis; Hib.
XX	
OS	Haemophilus.

[illegible]


```

1285 aThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 1301
1295 GCTACTAAG.....GATCAGCGTCTCTACTCAAG 2003
1302 AlaThrGluGlyAlaAlaThrLeuThrSerSerGlyLysLeuThrTh 1318
2004 GTATTTCACAACTTATGACACGAATGTAGATACAGTGTAAAAGTGGGG 2053
1318 rGluAlaSerSerHisIleThrSerAlaLysGlyGlnValAsnLeuSerA 1335
2054 CT.....CTGGGAGGAGTTAAACGACGACGACGAGAGTG 2088
1335 laglnaspGlySerValAlaGlySerIleAsnAlaAla.....AsnVal 1349
2089 ATACCCACAGCAGAGTGGAGCACTGTACATACCTGGCTGGATTGAGAATGA 2138
1350 ThrLeuAsnThrThrGlyThrLeuThr.....ValLysGlyse 1363
2139 TGAATACATGGAATCCACACAGACCTGAATTAATAAGATGATGTTTC 2188
1363 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 1380
2189 AA.....CACAAAGCAAGTGTCTTTCAGCAGA 2214
1380 luleAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn 1396
2215 ACATCTCGGAGGCTCAATTGGCTCTGTGATCTCCCAATGCTCCCAT 2264
1397 AlaAsnGlySerGlySerValIleAlaThrSerSerArgValAsnII 1413
2265 ACCT...GATCTCTCCCACTGGCCAAATC.....ACCG 2296
1413 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleIleSerLysAsnG 1430
2297 ACCTGAAGCGGGAATTCACGGGGGAGTCTCATTAATCTGACTTGACA 2346
1430 lyleAsnThrValLeuLeuLysGlyValLysIleAspValLysTyrIle 1446
2347 GCTCTCTGGG...GATGATATGACCATGGAACAGCTCACAAATATATCAT 2393
1447 GluProGlyIleAlaSerValAspGluValIleGluAlaLys..... 1460
2394 TCGAATAAGTACAGTATCTTGTATCTCAGACACAGTTCATCAATCTC 2443
1461 .ArgIleLeuGluLysValLysAspLeuSerAspGluGluArgGluAlaL 1477
2444 TT 2445
1477 eu 1477

```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA01846

seq_documentation_block:

ID AA01846 standard; Protein; 1536 AA.

XX AA01846;

XX 11-SEP-2000 (first entry)

DE Haemophilus influenzae strain 12 HMW1A protein, SEQ ID NO:67.

KW HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.

XX Haemophilus influenzae strain 12.

OS WO200020609-A2.

XX 13-APR-2000.

PD

```

XX 07-OCT-1999; 99WO-CA00938.
XX 07-OCT-1998; 98US-0167568.
XX 08-DEC-1998; 98US-0206942.
XX (CONN-) CONNAUGHT LAB LTD.
XX Loosmore SM, Yang Y, Klein MH;
XX WPI; 2000-303789/26.
XX N-PSDB; AAA52195.
XX Nucleic acid molecule for producing recombinant high molecular weight
XX proteins of Haemophilus which are used as a vaccine to provide
XX protection against Haemophilus induced diseases in humans -
XX Example 16; Fig 28A-Q; 307pp; English.
XX
XX The invention relates to the recombinant production of Haemophilus
XX influenzae high molecular weight (HMW) proteins in Escherichia coli. The
XX expression construct used to effect recombinant expression comprises a
XX promoter functional in E. coli (e.g., the T7 promoter) operably linked
XX to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
XX influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
XX clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA,
XX hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins
XX and the hmwB and hmwC genes encode accessory proteins which are
XX responsible for post-translational processing and secretion of the HMW
XX proteins. The modified hmwABC operon used in the expression construct of
XX the invention contains an A gene modified such that it encodes only the
XX mature HMW. The invention also discloses hmwA genes (AAA52175-A52198)
XX and HMW proteins (AAB01824-B01849) from the non-typeable H. influenzae
XX strains Joys, K1, K21, LCD2, PMH1, 15 and 12. The nucleic acids and
XX vectors are used for the production of recombinant H. influenzae HMW
XX proteins which can be used as vaccines to mediate a humoral or
XX cell-mediated immune response to provide protection against diseases in
XX humans caused by H. influenzae (e.g., otitis media, epiglottitis,
XX pneumonia and tracheobronchitis). The HMW proteins are also useful as
XX antigens in immunoassays for detecting antibodies against Haemophilus,
XX HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
XX HMW proteins can be used to isolate and clone hmw genes from other
XX non-typeable strains of Haemophilus via hybridisation reactions. The
XX present sequence represents an HMW protein from a non-typeable strain of
XX H. influenzae.
XX Sequence 1536 AA;

```

alignment_scores:
Quality: 138.50 Length: 936
Ratio: 0.318 Gaps: 49
Percent Similarity: 46.581 Percent Identity: 20.406

alignment_block:

US-09-049-696-18 x AAB01846 ..

Align seg 1/1 to: AAB01846 from: 1 to: 1536

```

217 TTGTAAGCTACA.....GGAAGCGATTTTATTTCAAAATGT 254
631 PheGluGlyThrLeuAsnIleSerGlyLys.....ValAsnII 643
255 TGCATTTTGTATTCCTGAAACATGGAAGCTGACTGATGTGAGAC 304
643 eSerMetValLeuPro.....LysAsnGluSerGlyTyrAspLysP 657
305 CAAACTTGAGACCTACAAAATGCTGATGTTCTG...GTTGCTGAGTCT 351
657 heLysGlyArgThrTyrTrpAsnLeuThrSerLeuAsnValSerGluSer 673
352 .....ACTCTCCAGGTATGAT..... 369

```

674 GlyGluPheAsnLeuThrIleAspSerArgGlySerSerAlaGlyTh 690
370GAACCCACACT.....GAGCAGA 388
690 rLeuThrGlnProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThr 707
389 TGGCAACTGTGGAGAGA.....AGGTGAAGATCCACTCCT 432
707 hrPheAsn.ValGluArgAsnAlaArgValAsnPheAspIleLysAlaPr 723
433 GATTTCATTGAGGAAAAAGTGTAGCTCAATATGACACCAAGGTAGGC 482
723 oileGlyIleAsnLysTyrSerSerLeuAsnTyrAlaSerPheAsnGlyA 740
483 ATTTGTCATGAGTGGGCTCATCTACGATGGGAGTATTTGACGAGTACA 532
740 snileSerValSerGly.....GlyGlySerValAspPheThr 752
533 ATATGATGAGAAATCTACTATCCATGGAAGAATACA.....AGCA 576
753 LeuLeuAlaSerSerSerAsn.ValGlnThrProGlyValValIleAsnS 769
577 GTAGATCTTCAGCAGGATTTACTGGTACAAATGTAGTAAAGAAGTGCA 626
769 erLysTyrPheAsn.ValSerThrGlySerSerLeuArgPheLysThrSe 785
627 GGA.....GCCACTGTTACACCACCAAGATGCACATCAAT. 663
785 rGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeuAsnA 802
664AAGTACAGGACTCTATGAAAGAGGATGTAGTTGTT 702
802 laThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspGlyMet 818
703 CTCCAATCCCGCAGGAGGAGGCTTCTATAATGTTT..... 741
819 ileGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyGlyAs 835
742GCACAACATGTTGATTTCTATAGTTG 766
835 nileThrPheGlySerArgLysAlaValThrGluIleGluGlyAsnVal 852
767 AATTCTGTACAGAA.....CAAAAC 786
852 hrIleAsnAsnAlaAsnValThrLeuIleGlySerAspPheAspAsn 868
787 CACAACAAGAGTCCAAACAGCAAAAT.....CAAAATGCAA 827
869 HisGlnLysProLeuThrIleLysLysAspValIleAsnSerGlyAs 885
828 TCTCCGAGCACATGGGAAGTG.....ATCC 853
885 nLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuThrValG 902
854 GTGATCTCGAGACTTTAAGAAACCACTCTCATGACA..... 891
902 luSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsnValGly 918
892ACACACCCACCAATCCCACTCTCATTTGTCGACATGG 932
919 GlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLysGlyG 935
933 ACAAGAATGTGTGTAGTCTTACAAATCTGGAAGCATGGGACTG 982
935 yalaArgPheLysAspIleAspAsnSerLysAsnLeuSerIleThrThra 952
983 GTAACCGCTC.....AATCGACTGAATCAA 1008
952 snSerSerThrTyrArgThrIleIleSerGlyAsnIleThrAsnLys 968
1009 GCAGCCACAGTTTCTCTCCACAGACTTGCAGTGGGGTCTGGGTGG 1058
969 AsnGlyAspLeuAsnIleThrAsnGlu.....GlySer..... 979

1059 GATGGTGACATTTGACAGTGCTGCCCATGTACAAAGTGAACACTCATACAG. 1107
980AspThrGluMetGlnIleGlyGlyAspValSerGlnL 992
1108ATAACAGTGGCAGTGACAGGGACACACTGCCCAAGA 1146
992 ysGluGlyAsnLeuThrIleSerSerAspLysIleAsnIleThrLysGln 1008
1147TTACCTGCAGCAGCTTCAGAGGAGGAGCTCCATCTGCACGGGCT 1190
1009 ileThrIleLysAlaGlyValAspGlyGluAsnSerAspSerAlaTh 1025
1191 TCGATCGGCTTTACTGTGATTAGGAAGAA..... 1221
1025 rAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrGlnA 1042
1222TATCCAACTGATGGA 1236
1042 spLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGly 1058
1237 TCTGAAATTTGCTGCTCAGCGATGGGAGACACACACTATAAGTGGG... 1284
1059 SerAsp.....LeuThrIleGlyAsnThrAsnSerAlaAspGlyTh 1072
1285TGCTTTACGAGGTCAACAAAGT.....GGTG 1312
1072 rAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 1089
1313 CCATCATCCACACAGCTGCTTTGGGCCCTCTGCAGCTCAAGAAGTAGAG 1362
1089 laAspGlyHisLysValThrLeuHis..... 1097
1363 GAGCTGTCCTCAAAATG...ACAGGAGTTTACAGACATATGCTTCAGATCA 1409
1098SerLysValGluThrSerGlySerAsnAsnAsnThrGluAspse 1112
1410 AGTTCAGACAATGSCCTCATTTGCTTTTGGGCCCTTTTCATCAGGAA 1459
1112 rSerAspAsnAsnAla..... 1117
1460 ATGGAGCTGCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGGATTAAAC 1509
1118GlyLeuThrIleAspAlaLysAsnValThr 1127
1510 CTCAGAACACCCAGCTGATGAATGACAGTATGCTGGACACACCGT 1559
1128 ValAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSe 1144
1560 GGGAAAGGACACTTTGTTCTTATCAGCTGGACACAGCAGCTCCCAAA 1609
1144 rGlyGlu.....IleThrThrLysThr..... 1151
1610 TCCTTCTCTGGATCCAGTGGCAGAGCAAGGTGGCTTTGTAGTGGAC 1659
1152GlyThrThrIleAsn 1156
1660 AAAACACCAAAATGGCTTACCTCAATCCAGGCATTCGTAAGTTGG 1709
1157 AlaThrThrGly.....AsnValGluIleThrAlaGlnThrGl 1169
1710 CACTTGGAAATACAGTCTGCAAGCAAGCTCACAAACCTTGACCTGACTG 1759
1169 yserIleLeuGlyGlyIleGluSerSerGlySerValThrLeuThra 1186
1760 TCAGTCCCGTGGG.....TCCAAATGCTGCTGCCTCCCAATTACA 1800
1186 laThrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThr 1202
1801 GTGACTTCCAAACGACCAAGACACACAGC..... 1830
1203 ValThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLy 1219

```
1830 ..... 1830
1219 sGlyThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyT 1236
1831 .....AAATTCCCGAGCCCTCGGTAGTT 1854
1236 hrIleSerGlyGlyThrValGluValLysAlaThrGluSerLeuThrThr 1252
1855 TATGCAAAATATTCG .....CAAGGAGCCTCCCAATTCACG 1892
1253 GlnSerAsnSerLysIleLysAlaThrThrGlyGluAlaAsnValThrSe 1269
1893 GGCAGGTGCACAGCCCTGATTGAATCAGTGAATGGAAACACAGTTACCT 1942
1269 rAlaThrGlyThr...IleGlyGlyThrIleSerGlyAsnThrValAsnV 1285
1943 TGGAA.....CTACTGGATAATGAGCAGGTGCTGAT 1974
1285 alThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 1301
1975 GCTACTAAG.....GATGACGGTCTCTACTCAAG 2003
1302 AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 1318
2004 GTATTTTACACACTTATGACACAGATGTAGATACAGTGTAAAGTGGGG 2053
1318 rGluAlaSerSerHisIleThrSerAlaLysGlyGlnValAsnLeuSerA 1335
2054 CT.....CTGGGAGGAGTTAACCGACCGCAGCAGGAGTG 2088
1335 laGlnAspGlySerValAlaGlySerIleAsnAlaAla.....AsnVal 1349
2089 ATACCCAGCAGAGTGCAGACACTGTACATACCTGGCTGGATTGAGATGA 2138
1350 ThrLeuAsnThrThrGlyThrLeuThrThr.....ValLysGlySe 1363
2139 TGAATAACAATGGAATCCACCAAGACCTGAAATTAATAGGATGATGTC 2188
1363 rAsnIleAsnAlaThrSerGlyThrLeuValIleAsnAlaLysAspAlaG 1380
2189 AA.....CACAAGCAAGTGTCTTTCACGAGA 2214
1380 luLeuAsnGlyAlaAlaLeuGlyAsnHisThrValValAsnAlaThrAsn 1396
2215 ACATCTCGGAGGAGCCTATTGTTGGCTTCTGATGTCCTCCCAATGCTCCAT 2264
1397 AlaAsnGlySerGlySerValIleAlaThrThrSerSerArgValAsnII 1413
2265 ACCT...GATCTCTTCCACCTGGCCAAATC.....ACCG 2296
1413 eThrGlyAspLeuIleThrIleAsnGlyLeuAsnIleSerLysAsnG 1430
2297 ACCTGAAGCGGAAATTCACGGGGCAGTCTCATTAATCTGACTTGACA 2346
1430 lyIleAsnThrValLeuLeuLysValLysIleAspValLysTyIle 1446
2347 GCTCCCTGGG...GATGATTATGACCATGGAGACCTCACAAATATATCAT 2393
1447 GlnProGlyIleAlaSerValAspGluValIleGluAlaLys..... 1460
2394 TCGAATAAGTACAGTATTCTTGATCTCAGAGACAGTTCATGAATCTC 2443
1461 .ArgIleLeuGluLysValLysAspLeuSerAspGluGluArgGluAlaL 1477
2444 TT 2445
1477 eu 1477
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA89860

seq_documentation_block:

ID AA89860 standard; Protein; 1152 AA.

XX

```
AC AA89860;
XX
DT 26-SEP-2001 (first entry)
XX
DE C-glutamicum protein fragment SEQ ID NO: 3614.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS CcFynebacterium glutamicum.
XX
PN EF1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
XX
PR 07-APR-2000; 2000JP-0159162.
XX
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOM ) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH65079.
XX
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX
XX Claim 17; SEQ ID NO: 3614; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from Coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a protein described
XX in the exemplification of the invention.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX European Patent Office.
XX
SQ Sequence 1152 AA;
```

```
alignment_scores:
  Quality: 135.50      Length: 801
  Ratio: 0.383        Gaps: 38
  Percent Similarity: 44.195    Percent Identity: 19.476
```

alignment_block:

US-09-049-696-18 x AA89860 ..

Align seg 1/1 to: AA89860 from: 1 to: 1152

```
55 TTGATTCTTCACCTTCTAGAGGGGCCCTGAGTAATTCACCTCATTCAGCT 104
||||| :||| :||| :||| :||| :|||
478 LeuValProPheIleAsnGluProAspThrAsnProAlaLeuIleGlnAs 494
-
105 GACACACATGGCTATGAAGGCATTGCTGTGCAATGACCCCAATGTC 154
||||| :||| :||| :||| :||| :|||
494 nGlnGlnTrpAlaThrLysAlaValAlaThrAlaAlaGluProGlyTrpL 511
155 CAGAAGATGAACA..... 168
||| :|||
```


PR 08-JUL-1999; 99DE-1031424.

SQ Sequence 1152 AA;

alignment_scores:

Quality:	135.50	Length:	801
Ratio:	0.383	Gaps:	38
Percent Similarity:	44.195	Percent Identity:	19.476

alignment_block:
 US-09-049-696-18 x AAB79787 ..

Align seg 1/1 to: AAB79787 from: 1 to: 1152

55 TTGATTCTTCACCTCTCTAGAGGGGCCCTGAGTAATTCACCTCATTCACGCT 104
 |||::: :::: ||| :::: :::: |||::|
 478 LeuValProPheIleAsnGluProAspThrAsnProAlaLeuIleGlnAs 494
 105 GAACACAATGGCTATGAGGCATTTGCTGTGCAATCGACCCCATGTGTC 154
 ::::: ::::: ::::: ::::: ||| ::::: |||::|
 494 nGlnGlnTrpAlaThrLysAlaValAlaThrAlaAlaGluProGlyTrpL 511
 155 CAGAAGATGAACA..... 168
 ||| ::::|
 511 euGluLysGlnThrLysProGluValLeuGluGluYAspValAspLys 527
 169 CTCATTCAACAATAAGGACATGGTCACCCAGGCATCTCTGTATCTGTT 218
 |||::|::: ::::: ||| ::::: |||
 528 LeuIleAsnAspValArgAspAlaAla..... 536
 219 TGAAGCTACAGGAAGCGATTTTATTTTCAAAATGTTGCCATTTTGATTC 268
 536 536
 269 CTGAACATGGAAGACAAAGGCTGACTATGTGAGACAAACATTGAGACC 318
 |||::|::: ::::: ||| ::::: |||
 537 ..GluAlaTrpAlaAlaArgProAlaAlaArgGluAlaGluIle...Leu 551
 319 TACAAAATGCTGATGTTCTGGTGTCTGAGTCTACTCTCCAGGTAATGA 368
 |||::|::: |||::: ||| ::::: |||
 552 TyrLysThrAlaGluIleLeuArgVal..... 560
 369 TGAACCTTACACTGAGCAGATGGCAACTGTGGAGAGAAGGTTGAAAGGA 418
 561ArgArgGlyHisLeuI 566
 419 TCACCTCTACTCTGATTCATTCACGAGAAAGTTAGCTGAATATGGA 468
 ||::|::| ::::: |||::| ::::: |||
 566 leSerValThrAlaAlaGluVal...GlyLysAlaValGluGlnThrAsp 581
 469 CCA.....CAAGTAGGGCATTTGTCCATGAGTGGGCTCATCTACG 509
 ||| ::::: |||::: |||::: |||::: |||
 582 ProGluIleSerGluAlaIleAspPheAlaArgTyrAlaHisLeuAl 598
 510 ATGGGAGATTTTGAC.....GAGTACAATAATGATGAGAAT 547
 ::::: ||| |||::: |||::: |||
 598 aLeuGluLeuAspValAspAlaGluPheThrProAspArgVal 615
 548 TCTACTTATCCAATGGAGAATACAAGCAGTAAGATGTTTCACAGCGTATT 597
 ::::: ||| ::::: |||
 615 alValValThrProTrpAsnPheProIleAlaIleProAlaGlySer 631
 598 ACT.....GGTACAANTGTAGTAAGAAGTGTCCAGG 629
 * ||| |||::: |||::: |||
 632 ThrPheAlaAlaLeuAlaAlaGlyValIleHisLys..... 645
 630 AGGCAGCTGTTACACCAAAAGATGCACATTCATTAAGTCAACAGACTCT 679
 ||| ::::: |||::: |||
 646 .ProSerLysProSerGlnHisCysSerAlaAlaValValGluAlaLeuT 662
 680 ATGAAAAGGA.....TGAGTGTGTTCTCCAATCCCGC 714
 ::|::| ||| ::::: |||
 662 rpGluAlaGlyValProArgGluValLeuHisCysIleTyrProAlaAsn 678
 ::::: ||| ::::: |||

```
1919 ...GlyTrpLysSerSerValGlyLeuGlySerLysAlaGlyGlyPr 934
1569 CACTTTCTTTCTTATC.....ACCTGGACACGCGCCT.....1602
...:.....:.....:.....:.....:.....:.....:.....:.....:
934 oAsnTyrrValMetLeuMetGlyThrTrpAlaAspAlaProSerHisHis 951
...:.....:.....:.....:.....:.....:.....:.....:.....:
1603 ..CCCAAAATCTCTCTGGATCCCAAGTCGACAGAGCAAGGTGGCTTT 1650
...:.....:.....:.....:.....:.....:.....:.....:.....:
951 laProArg.....:.....:.....:.....:.....:.....:.....: 953
1651 GTAGTGGCAAAACACCAAAATGGCTACCTCAAAATCCAGCGCATTCG 1700
...:.....:.....:.....:.....:.....:.....:.....:.....:
954 .....GluThrAsnProLeuLysSerLysLeuAspLeuProGly..... 966
1701 TAAGGTGGCACTTGGAAATACACTCTGCAAGCAAGCTCACAAACCTTGA 1750
...:.....:.....:.....:.....:.....:.....:.....:.....:
967 .GluGluLeuGluTrpLeuGlyLysAlaAsnAlaSerAspGluThr..... 981
1751 CCCTGACTGTACAGTCCGCGTCCCAATGCTACCCCTGCCCTCCAAATACA 1800
981 .....:.....:.....:.....:.....:.....:.....:..... 981
1801 GTGACTTCCAAAGCAAGACACGACCAAAATTCGCCAGCCCTCTGGT 1850
982 ...AlaTrpAsnThrGluPheGlySerProArgAspProSerGlyLeuAs 997
1851 AGTTTATGCAAAATTT.....:.....:.....:.....:..... 1866
...:.....:.....:.....:.....:.....:.....:.....:.....
997 pValGluAlaAsnIlePheArgTyrrArgProAlaGluValValLeuArgL 1014
1867 .....:.....:.....:.....:.....:.....:.....:..... 1890
1014 euAspAspSerAlaThrProArgGluThrAlaArgAlaLeuLeuAlaLa 1030
1891 ...AGGCCAGTGTACAGCCCTGATTGAATCAGTGAATGGAACACAGT 1937
...:.....:.....:.....:.....:.....:.....:.....:.....
1031 ArgArgAlaGlyValThrProArgValLeuGluThrProGlyValSerGl 1047
1938 TACCTTGAAGTGTGTAATGAGCAGGTGTGTGATGCT...ACTAAGG 1984
...:.....:.....:.....:.....:.....:.....:.....:.....
1047 uGlnValArgGluValLeuSerAlaAlaGlyValSerAlaGluThrVala 1064
1985 ATCAGCGTGTCTACTCAAGTATTTCACA.....ACTTATGACACCAAT 2028
...:.....:.....:.....:.....:.....:.....:.....:.....
1064 spAspSerValPheIleSerAsnValLeuArgGlyGluTyrrAspGluAsn 1080
2029 GGTAGATACAGTGTAAAGTGCAGGCTCTGGGAGGAGTTAAC..... 2070
...:.....:.....:.....:.....:.....:.....:.....:.....
1081 Ser.....SerValArgValArgTyrrLeuGlyLysValSerAspThrVa 1095
2071 .GCAGCCAGACGGAGAGTATACCCAG.....:.....:.....:..... 2097
...:.....:.....:.....:.....:.....:.....:.....:.....
1095 laArgGluArgLeuSerValArgProGluValValLeuLeuAspAlav 1112
2098 .....CAGAGTGGCAGCTACATCTGCTGGTGGATGAGATGATGAA 2142
1112 alThrAlaSerGlyArgValGluLeuArgTyrrTrpLeuLysGluGlnAla 1128
2143 ATA 2145
1129 Ile 1129
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAU05396

seq_documentation_block:

ID AAU05396 standard; Protein; 26926 AA.

XX AC AAU05396;

XX DT 24-OCT-2001 (first entry)

XX DE Human titin (connectin) protein sequence.

```
XX Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;
KW titin-related disease; zebrafish; heart failure; heart disease.
OS Homo sapiens.
XX NC_020015.1:66-A1.
XX 19 JUL-2001.
XX 12-JAN-2001; 2001WO-US01212.
XX 12-JAN-2000; 2000US-0175787.
XX (GEO) GEN HOSPITAL CORP.
XX Fishman MC;
XX WPI; 2001-451869/48.
XX N-PSDB; AAS05390.
XX Determining if a subject has or is at risk of developing a
XX titin-related disease or condition, particularly heart failures,
XX comprises detecting the presence of a mutation in the titin gene
XX Disclosure: Page 57-111; 114pp; English.
XX The present sequence representing human titin (also known as connectin)
XX is described in an invention relating to a novel method for determining
XX whether a subject has or is at risk of developing a titin-related
XX disease or condition. The method comprises analysing a nucleic acid of
XX a sample from the subject and detecting the presence of a mutation
XX (e.g. the pickwick mutation in the cardiac specific exon N2B) in the
XX titin gene, which indicates that the subject has or is at risk of
XX developing a titin-related disease. The zebrafish which has a phenotype
XX similar to mammalian heart failure is used as a model. The method is
XX useful for detecting an increased likelihood of heart disease, such as
XX heart failure, in a patient, so that appropriate intervention can be
XX instituted before any symptoms occur. The method may also be used to
XX facilitate determination of etiology of an existing heart condition,
XX such as heart failure, to identify compounds that can be used to treat
XX or prevent heart conditions, in prenatal genetic screening, e.g. to
XX identify parents who may be carriers of a recessive titin mutation.
XX Compounds identified using the methods may be used to treat patients
XX that have or are at risk of developing heart disease, e.g. heart
XX failure.
```

SQ Sequence 26926 AA;

alignment_scores:

Quality:	134.00	Length:	851
Ratio:	0.398	Gaps:	40
Percent similarity:	39.600	Percent identity:	18.801

alignment_block:

US-09-049-696-18 x AAU05396 ..

Align seg 1/1 to: AAU05396 from: 1 to: 26926

```
610 GTAGTAAAGAGTGTACGAGGAGCGCTGTACACCAAAAGATGCACATT 659
660 ILeileGluArgCysGluGluGlyLysAspAsnTrpIleArgCysAsnMe 6846
660 CAAT.....AAAGTAACAGGACTCTATGAAAAAG 688
.....:.....:.....:.....:.....:.....:.....:.....
6846 tLysLeuValProGluLeuThrTyrrLysValThrGlyLeu....GluLysG 6862
689 GATGTGAGTTTGTCTCCAAATCCGCCAGACGAGGAGGCTTCTATAATG 738
|| .....:.....:.....:.....:.....:.....:.....:.....
6862 lyAsnLysTyrrLeuTyrrArgValSerAlaGluAsnLysAlaGlyValSer 6878
739 TTTGCACACATGTT.....:.....:.....:.....:.....:..... 767
```

```

6879 AspProSerGluIleLeuGlyProLeuThrAlaAspAlaPheValG1 6895
      ::::: |||:::
768 ATTCTGTACAGACAAACCAACAAAGAGCTCCAAACAGCAAAATC 817
      ::::: |||:::
6895 uProThrMetAspLeuSerAlaPheLysAspGly 6906
      ::::: |||:::
818 AAAATGCNATCTCGAAGCACATCGGGAAGTATCGGTGTTCTGAGGAC 867
      ::::: |||:::
6907 6907 .....LeuGluValIle 6910
868 TTTAAGAAACCACTCTATGACACACAGCCACCAAAATCCACCTTCTC 917
6911 .....ValProAsnPro 6914
918 ATTGCTGCAGATTGGACAAAGAATTGTGTAGTCTTGACAAATCTG 967
      ::::: |||:::
6915 .....IleThrIleLeuValProSerThrGlyT 6924
968 GAAGCATGGGACTGGTAACCCCTCAATCGACTGAATCAGCAGCGCCAG 1017
      ::::: |||:::
6924 yrProArgProThr 6931
1018 CTTTCTCTGCTGCACAGATTGAGCTGGGTCTGGTGGGTGAGTAC 1067
      ::::: |||:::
6932 CysPheGlyAspLysValIleGluThrGlyAspArgValLysMetLysTh 6948
1068 APTTGACAGTGTGCCATGACAAAGTGAATCATACAGATAAAGACAGTG 1117
      ::::: |||:::
6948 rLeuSerAlaTyrAlaGluLeu 6962
      ::::: |||:::
1118 GCAGTGACAGGACACACTCGCCAAAGATTACCTGCAGCAGCTTCAGGA 1167
      ::::: |||:::
6962 rGSerAspLysGlyIleTyrThrLeuLysLeuGlu 6973
1168 GGGACGTCCATCTGCAGCGGGCTTCGATCGGCATTTACTGTGATTAGGAA 1217
6973 6973 ..... 6973
1218 GAAATATCCAACTGATGATCTGAAATTTGCTGCTGACGATGGGAAG 1267
6973 6973 ..... 6973
1268 ACAACACTATAGTGGTGCCTTACGAGGTCAACAA...AGTGGTCCC 1314
6974 .....AsnArgValLysThrIleSerGlyGlu 6982
1315 ATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAG... 1362
6983 IleAspValAsnValIleAlaArgProSerAlaProLysGluLeuLysPh 6999
1363 ....GAGCTGTCCTCAAAATGACAGAGGAGTTTACAGACATATGCTTCAGATC 1408
6999 eGlyAspIleThrLysAspSerValHisLeuThrTrpGluProProAsp 7015
1409 ANGTTCAGACAAATGGCCTCATGTATGCTTTTGGGGCCCTTTCATCAGGA 1458
7016 .....AspAspGlyGlySerProLeuThrGly 7024
1459 AATGAGAGTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAC 1508
7025 Tyr...ValValGluLysArgGluValSerArgLysThrTrpThrLysVa 7040
1509 CTTCCAGAACCCAGTGGATGAATGGACAGATGATCGTGGACACACCG 1558
7040 lMetAspPheValThrAspLeuGluPheThrVal...ProAspLeuValG 7056
1559 TGGGAAGGACACTTTGTTCTTATC 1584
7056 lngLysGluTyrLeuPheLysValCysAlaArgAsnLysCysGlyPro 7072
1585 .....ACCTG 1589

```

```

7073 GlyGluProAlaTyrValAspGluProValAsnMetSerThrProAlaTh 7089
1590 GACAACGAGCTCCCAA... 1608
7089 rValProAspProProGluAsnValLysTrpArgAspArgThrAlaAsnS 7106
1609 .....ATCCTTCTCTGGGATCCC...AGTGGACAGAGCAA 1641
7106 erIlePheLeuThrTrpAspProProLysAsnAspGlyGlySerArgIle 7122
1642 GGTGGCTTTGTAGTGACAAA... 1662
7123 LysGlyTyrIleValGluArgCysProArgGlySerAspLysTrpValAl 7139
1663 .....AACACAAATGGCTTACCTCCAAATCCAG 1693
7139 aCysGlyGluProValAlaGluThrLysMet...GluValThrG 7153
1694 GCATTGCTAAGTTGGCACTGG...ANATACAGT 1725
7153 lyLeuGluGlu...GlyLysTrpTyrAlaTyrArgValLysThrLeuAsn 7168
1726 CTGCAAGCAAGCTCACA... 1743
7169 ArgGlnGlyAlaSerLysProSerArgProThrGluGluIleGlnAlaVa 7185
1744 .....ACCTTGACCCCTGA 1756
7185 lAspThrGlnGluAlaProGluIlePheLeuAspValLysLeuLeuAlaG 7202
1757 CTGTCACGTCGGTCCCAATGCTACCTCCCTCCCAATACAGTACTG 1806
7202 lyLeuThrValLysAlaGlyThrLysIleGluLeuProAlaThrValThr 7218
1807 TCCAAACCAACAGCAGCAGCAAAATCCCGCCCTCTGGTAGTTTA 1856
7213 GlyLys...ProGluProLysIleThrTr 7227
1857 TGCAAAATTCGCCAAGGAGCTCCCAATTCACGGCCAGTGTACAG 1906
7227 pThrLys...AlaAspMetIleLeuLysGlnAspLysArgI 7240
1907 CCTGATTGAATCAGTGAATGGA...ACAGTTACCTTGGAACTACTG 1953
7249 leThrIleGluAsnValProLysLysSerThrValThrIle... 7253
1954 GATAATGGAGCAGGTGCTGCTACTACTAAGGATGACGGTGTCTACTCAAG 2003
7254 .....ValAspSerLysArgSerAsp... 7260
2004 GTATTTCACAATTATGACAGCAATGGTAGATACAGTCTAAAAGTGGCGG 2053
7261 .....ThrGlyThrTyrIleIleGlu... 7267
2054 CTCGGGAGGAGTTAACGCGCAGCAGAGAGTATACCCAGCAGAGT 2103
7268 .....AlaValAsnValCysGlyArgAlaThrAlaValValGluVal 7281
2104 GGAGCACTGTATACCTGGC...TGGAT 2129
7282 AsnValLeuAspLysProGlyProProAlaAlaPheAspIleThrAspVa 7298
2130 TGAGAATGATGAA...ATACATGGAATCCCAAGACCTGAA... 2169
7298 lThrAsnGluSerCysLeuLeuThrTrpAsnProProArgAspGlyG 7315
2170 .....ATTAATAAGGATCAT 2184
7315 lySerLysIleThrAsnTyrValValGluArgAlaThrAspSerGlu 7331
2185 GTTCAACACAGCAAGTGTGTTTCAGCAGAAACATCCTCGGAGGCTCAT 2234
7332 ValTrpHisLys...LeuSerSerThrValLysAspThrAsnPh 7345

```


2235 TGTGGCTTCTGAT...GTCCCAAT..... 2256
 | |||::: ::|||
 7345 eLysAlaThrLysLeuIleProAsnLysGluTyrIlePheArgValAlaA 7362
 2257GTCCTCATACCT 2268
 7362 laGluAsnMetTyrGlyAlaGlyGluProValGlnAlaSerProIleThr 7378
 2269 GATCTTTC.....CCACTGCCAAATCACCGACTGAAGCGGA 2309
 :: ||||| ||| |||::: ::
 7379 AlaLysTyrGlnPheAspProGlyProProThrArgLeuGluProSe 7395
 2310 AATTCACGGGGCAGTCTATTATCTGACTGGACAGCTCCTGGGGATG 2359
 : ::: ::||| ||| ||| |||
 7395 rAspIleThrLysAspAlaValThrLeuThrTrpCysGluProAspAspa 7412
 2360 ATTATGACCATGGAACAGCTCACAAGTATATCATTCGAAATAGTACAAGT 2409
 || ::: ::||| |||::: ::
 7412 spGlyGlySerProIleThrGlyTyrTrpValGluArgLeuAspProAsp 7428
 2410 ATTCTTGATCTCAGACAGTTC.....AATGAATCTCTTCAACTGAA 2453
 |||||::: ::|||::: |||
 7429ThrAspLysTrpValArgCysAsnLysMetProVally 7441
 2454 TACTACTGCTCTCATCCCAAGGAGCCAACTCTGAGGAGTCTTTTGT 2503
 : |||::: ||| ::|||::: ::|
 7441 sasPThrThrTyrArgValLysGlyLeuThrAsnLysLysLysTyrArgp 7458
 2504 TTAACACAGAAACATTACTTTTGAATAATGGCACAGATCTTTTCATTGCT 2553
 ||::: ::||| |||::: ::
 7458 heArg.....ValLeuAlaGluAsnLeuAlaGlyProGlyLysPro 7471
 2554 ATTCAGGCTGTTGATAAGGTCGATCTGAATCAGAAATATCCAAACATTGC 2603
 : |||::: ::|||::: |||::: |||
 7472 SerLysSerThrGluProIleLeuIleLysAspProIleAsp..... 7485
 2604 ACGAGTATCTTTGTTTATTCCTCCACAGACTCCGCCAGACACACTAGTC 2653
 ||| ||||| |||::: |||
 7486ProProTrpProProGlyLysProThrV 7495
 2654 CTGATGAACGCTGCTCTGCTTGTCTTAATATTCATATCAACAGCACCATT 2703
 : |||::: ::|||::: ||| |||
 7495 alLysAspValGlyLys.....ThrSerValArgLeuAsnTrpThrLys 7509
 2704 CCT 2706
 |||
 7510 Pro 7510

This Page Blank (uspio)

OM of: US-09-049-696-17 to: SPTREMBL_17:* out_format : pfs
Date: Mar 30, 2002 2:47 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cgnt2_1/USPTO_spool/US09049696/runat_28032002_145238_2053/app_query.fasta_1.12579
-DB-SPTREMBL_17 -QFMT=fastan -SUFFIX=rspt -CAPOP=12.000
-GAPEXT=4.000 -MINMAST=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-CGAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696_CGNTL_1_705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT THREADS=1

Search information block:

Query: US-09-049-696-17
Query length: 106
Database: SPTREMBL_17:*
Database sequences: 473505
Database length: 14622329
Search time (sec): 805.760000

score_list:

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation
sp_human:O95151	+	107.00	284.20	1.6e-08	914	1 O95151 homo sapiens (human) ...
sp_human:Q9UPC6	+	107.00	284.20	1.6e-08	914	1 Q9UPC6 homo sapiens (human) ...
sp_human:Q9UNF6	+	107.00	284.20	1.6e-08	914	1 Q9UNF6 homo sapiens (human) ...
sp_mammal:Q9TUB5	+	105.00	278.36	3.4e-08	917	1 Q9TUB5 sus scrofa (pig) ...
sp_rodent:Q88826	+	90.00	234.77	9.2e-06	913	1 Q88826 mus musculus (mouse) ...
sp_rodent:Q9D726	+	90.00	234.77	9.2e-06	913	1 Q9D726 mus musculus (mouse) ...
sp_invertebrate:Q96532	+	59.50	150.59	0.7875	520	1 Q96532 drosophila melanogaster ...
sp_plant:Q9C7B2	+	57.00	136.50	2.06	1211	1 Q9C7B2 arabidopsis thaliana (m ...
sp_plant:Q9LHK2	+	57.00	136.49	2.06	1213	1 Q9LHK2 arabidopsis thaliana (m ...
sp_invertebrate:Q9NJG3	+	52.00	152.65	11.66	27	1 Q9NJG3 steatoda triangulosa ...
sp_bacteria:Q9XAU2	+	52.00	136.53	12.50	199	1 Q9XAU2 pseudomonas stutzeri (ps ...
sp_rodent:Q9WU91	+	52.00	132.16	12.74	342	1 Q9WU91 mus musculus (mouse) ...
sp_plant:Q9ST49	+	52.00	126.17	13.07	719	1 Q9ST49 zea mays (maize) ...
sp_plant:Q94325	+	52.00	125.93	13.08	740	1 Q94325 arabidopsis thaliana (m ...
sp_invertebrate:Q9BI48	-	52.00	112.87	13.84	3734	1 Q9BI48 caenorhabditis elegans ...
sp_plant:Q9S131	+	51.50	125.96	15.66	616	1 Q9S131 arabidopsis thaliana (m ...
sp_human:Q9U002	+	51.00	117.21	19.48	1521	1 Q9U002 homo sapiens (human) ...
sp_human:Q9UPY3	+	51.00	115.31	19.64	1924	1 Q9UPY3 homo sapiens (human) ...
sp_invertebrate:Q9VCA1	+	50.50	125.11	22.55	477	1 Q9VCA1 drosophila melanogaster ...
sp_invertebrate:Q16680	-	50.50	123.79	22.67	562	1 Q16680 caenorhabditis elegans ...
sp_invertebrate:Q9GZ75	+	50.50	118.39	23.21	1098	1 Q9GZ75 plasmodium falciparum ...
sp_invertebrate:Q9GZ76	+	50.50	116.50	23.40	1387	1 Q9GZ76 plasmodium falciparum ...
sp_invertebrate:Q9BK45	+	50.50	109.93	24.07	3130	1 Q9BK45 plasmodium falciparum ...
sp_invertebrate:Q9BK45	+	50.50	109.62	24.11	3254	1 Q9BK45 plasmodium falciparum ...
sp_invertebrate:Q25873	+	50.00	147.45	24.51	25	1 Q25873 polycelis felina ...
sp_human:Q9H8F7	+	50.00	127.11	26.77	311	1 Q9H8F7 homo sapiens (human) ...
sp_invertebrate:Q18550	+	50.00	125.26	26.98	391	1 Q18550 caenorhabditis elegans ...
sp_human:Q9NV04	+	50.00	124.98	27.01	405	1 Q9NV04 homo sapiens (human) ...
sp_plant:Q24569	+	50.00	120.66	27.52	692	1 Q24569 zea mays (maize) ...
sp_virus:Q900C4	+	49.50	130.11	31.64	179	1 Q900C4 tt virus. orfl (fragment ...
sp_bacteria:Q9PRV6	-	49.50	120.94	32.92	558	1 Q9PRV6 chlamydia muridarum. hyp ...
sp_plant:Q9ST24	-	49.50	118.67	33.24	739	1 Q9ST24 arabidopsis thaliana (m ...
sp_invertebrate:Q9BIR5	+	49.50	112.29	34.17	1629	1 Q9BIR5 drosophila melanogaster ...
sp_invertebrate:Q9VT28	+	49.50	112.28	34.17	1632	1 Q9VT28 drosophila melanogaster ...
sp_invertebrate:Q9GPT8	+	49.50	106.17	35.09	3479	1 Q9GPT8 drosophila melanogaster ...
sp_invertebrate:Q26402	+	49.00	143.92	35.69	27	1 Q26402 stylaria lacustris. lab ...
sp_invertebrate:Q27342	+	49.00	143.92	35.69	27	1 Q27342 stylaria lacustris. lab ...
sp_invertebrate:Q9Y192	+	49.00	143.92	35.69	27	1 Q9Y192 priapulus caudatus. lab ...
sp_invertebrate:Q46127	+	49.00	129.37	38.01	164	1 Q46127 helobdella triserialis ...
sp_rodent:Q9D5K9	+	49.00	126.50	38.48	234	1 Q9D5K9 mus musculus (mouse) ...

seq_name: sp_human:O95151
seq_documentation_block:
ID O95151 PRELIMINARY; PRT; 914 AA.
AC O95151;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
PC TISSUE=SMALL INTESTINE;
RX MEDLINE=99047526; PubMed=9828122;
RA Griber A.D., Elbie R.C., Ji H.L., Schreier K.D., Fuller C.M.,
Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
of human CLCAL, the first human member of the family of Ca2+-activated
Cl- channel proteins.";
RL Genomics 54:200-214(1998).
DR EMBL; AF039400; AAC95428.1;
DR InterPro; IPR000131; ATPase_gamma.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWFA; 1.
SQ SEQUENCE 914 AA: 100174 MW: 7066440C46557FA3 CRC64;

alignment_scores:
Quality: 107.00 Length: 20
Ratio: 5.350 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-17 x O95151
Align seg 1/1 to: O95151 from: 1 to: 914
1 GGCATTACATTTTAAATATGAGAGTGGATAGGAACTGCAGCT 50
|||||
895 GlycylleuleulysileMetTrpLysTrpIleGlyleuGlnLeu 911

seq_name: sp_human:Q9UPC6

seq_documentation_block:
ID Q9UPC6 PRELIMINARY; PRT; 914 AA.
AC Q9UPC6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Eible R.C., Ji H.L., Schreur K.D., Fuller C.M.,
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
RT of human CLCA1, the first human member of the family of Ca2+-activated
RT CI-channel proteins.";
RL Genomics 54:200-214(1998).
DR EMBL: AF039401; AAC95429.1; -.
DR InterPro: IPR000131; ATPase_gamma.
DR InterPro: IPR002035; vWEA_gamma.
DR PROSITE: PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE: PS50234; vWEA: 1.
DR SMART: SM00327; vWA: 1.
SQ SEQUENCE 914 AA; 100160 MW; 7761362B66752AA3 CRC64;

alignment_scores:
  Quality: 107.00      Length: 20
  Ratio: 5.350        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-17 x Q9UNF6 ..
Align seg 1/1 to: Q9UPC6 from: 1 to: 914

1 GGCATTACATTTTAAAAATATGTGGAGTGGATAGGAGAACTGCAGCT 50
|||||
895 GlylleHisIleLeuLysIleMetTrpLysTrpIleGlyGluLeuGlnLe 911

51 GTCAATAGCC 60
|||||
911 uSerIleAla 914

seq_name: sp_human:Q9UNF6
seq_documentation_block:
ID Q9UNF6 PRELIMINARY; PRT; 914 AA.
AC Q9UNF6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 1.
GN CACCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE, AND COLON;
RX MEDLINE=99364503; PubMed=10437792;
RA Agnel M., Vermet T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
RT chloride channel (CaCC) family predominantly expressed in the
RT digestive tract and trachea.";
RL PERS Lett. 455:295-301(1999).
DR EMBL: AF127036; AAD25487.1; -.
DR InterPro: IPR000131; ATPase_gamma.
DR InterPro: IPR002035; vWEA.
DR SMART: SM00327; vWA: 1.
DR PROSITE: PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE: PS50234; vWEA: 1.
SQ SEQUENCE 914 AA; 100202 MW; B58B451D97887D6B CRC64;

alignment_scores:
  Quality: 107.00      Length: 20
  Ratio: 5.350        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-17 x Q9UNF6 ..

```

```

Align seg 1/1 to: Q9UNF6 from: 1 to: 914

1 GGCATTACATTTTAAAAATATGTGGAGTGGATAGGAGAACTGCAGCT 50
|||||
895 GlylleHisIleLeuLysIleMetTrpLysTrpIleGlyGluLeuGlnLe 911

51 GTCAATAGCC 60
|||||
911 uSerIleAla 914

seq_name: sp_mammal:Q9TUB5
seq_documentation_block:
ID Q9TUB5 PRELIMINARY; PRT; 917 AA.
AC Q9TUB5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE EPIHELIAL CHLORIDE CHANNEL PROTEIN.
GN AECC.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ILEAL MUCOSA;
RA Gaspar K.J., Gabriel S.E., Racette K.J., Forsyth G.W.;
RT "A cDNA involved in porcine exocrine chloride conductance.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF095584; AAF00077.1; -.
DR InterPro: IPR000131; ATPase_gamma.
DR InterPro: IPR002035; vWEA.
DR PROSITE: PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE: PS50234; vWEA: 1.
DR SMART: SM00327; vWA: 1.
SQ SEQUENCE 917 AA; 100735 MW; 649F41944F7EAD19 CRC64;

alignment_scores:
  Quality: 105.00      Length: 20
  Ratio: 5.250        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 95.000

alignment_block:
US-09-049-696-17 x Q9TUB5 ..
Align seg 1/1 to: Q9TUB5 from: 1 to: 917

1 GGCATTACATTTTAAAAATATGTGGAGTGGATAGGAGAACTGCAGCT 50
|||||
898 GlylleHisIleLeuLysIleMetTrpLysTrpLeuGlyGluLeuGlnLe 914

51 GTCAATAGCC 60
|||||
914 uSerIleAla 917

seq_name: sp_rodent:O88826
seq_documentation_block:
ID O88826 PRELIMINARY; PRT; 913 AA.
AC O88826;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GGB-5 PROTEIN.
GN CLCA3 OR GGB-5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```

```
RP SEQUENCE FROM N.A.
RC TISSUE-INTESTINE;
RA Komiya T., Tanigawa Y., Hirohashi S.;
RT "Cloning of a gene, gob-5, which is expressed in intestinal goblet
   cells in mice.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017156; BAA33743.1; -.
DR MGD; MGI:1346342; Clca3.
DR InterPro; IPR000131; ATPase_gamma.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
SQ SEQUENCE 913 AA; 100070 MW; A7FA2F9E1089806D CRC64;

alignment_scores:
  Quality: 90.00 Length: 19
  Ratio: 4.737 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 68.421

alignment_block:
  US-09-049-696-17 x 088826 ..
Align seg 1/1 to: 088826 from: 1 to: 913

1 GCATTCACATTTTAAATATGTGGAAGTGGATAGGAGAACTGCAGCT 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
892 GlyileHisValLeuLysileMetTrpLysTrpLeuGlyGluMetGlnVa 908

51 GTCAATA 57
908 lThrLeu 910

seq_name: sp_rodent:Q9D7Z6

seq_documentation_block:
ID Q9D7Z6 PRELIMINARY; PRT: 913 AA.
AC Q9D7Z6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CHLORIDE CHANNEL CALCIUM ACTIVATED 3.
GN CLCA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=STOMACH;
RX MEDLINE=21083660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
DR EMBL; AK008659; BAB25815.1; -.
```

```
DR MGD; MGI:1346342; Clca3.
DR InterPro; IPR000131; ATPase_gamma.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
SQ SEQUENCE 913 AA; 100111 MW; 491E8584B06D9A89 CRC64;

alignment_scores:
  Quality: 90.00 Length: 19
  Ratio: 4.737 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 68.421

alignment_block:
  US-09-049-696-17 x Q9D7Z6 ..
Align seg 1/1 to: Q9D7Z6 from: 1 to: 913

1 GCATTCACATTTTAAATATGTGGAAGTGGATAGGAGAACTGCAGCT 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
892 GlyileHisValLeuLysileMetTrpLysTrpLeuGlyGluMetGlnVa 908

51 GTCAATA 57
908 lThrLeu 910

seq_name: sp_invertebrate:O96532

seq_documentation_block:
ID O96532 PRELIMINARY; PRT: 520 AA.
AC O96532;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE RAD17 PROTEIN.
GN RAD17 OR CG7825.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
GX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouch J., Brokstein P., Bröttier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke K., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
```

RA	DeIseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,	
RA	De-Simone V., Choisein N., Artiguenave F., Robert C., Brottier P.,	
RA	Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,	
RA	Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,	
RA	Wurmback E., Drzonek H., Erfle H., Jordan N., Bangert S.,	
RA	Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,	
RA	Vezzzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,	
RA	Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,	
RA	Reichelt J., Scharfe M., Schoen O., Barques M., Terol J., Clement J.,	
RA	Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,	
RA	Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masny D.,	
RA	de Haan M., Maarse A.C., Alcaraz J. P., Cottet A., Casacuberta E.,	
RA	Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,	
RA	Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Meves H.-W.,	
RA	Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,	
RA	Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,	
RA	Cresay T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,	
RA	Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,	
RA	Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,	
RA	Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,	
RA	Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,	
RA	Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,	
RA	Nakayama S., Nakazaki N., Shippo S., Takeuchi C., Wada T.,	
RA	Watanabe A., Yamada M., Yasuda M., Tabata S.,	
RT	"Sequence and analysis of chromosome 3 of the plant Arabidopsis	
RT	thaliana.";	
RL	Nature 408:820-822(2000).	
RL	EM3L; AC069474; XAG51027.1; -.	
RD	HW	
RD	Hypothetical protein.	
SW	SEQUENCE 1211 AA; 133616 MW; B681184F68AB39A4 CRC64.	

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

alignment_scores:		
"	Quality:	57.00
"	Ratio:	5.182
percent_similarity:		
"	Similarity:	91.667
percent_identity:		
"	Identity:	75.000
"	Gaps:	0
"	Length:	12

alignment_block:

Align seq 1/1 to: Q9C7B2 from: 1 to: 1211

25 TGGAGTGGATAGGAGAACTGCAGCTGTCAATAGCC 60

25 TGGAAAGTGGATAGGAGAACTGCAGCTGTCAATAGCC 60
|||:::|||||
267 TGAAGATTAValGlyLeuGlnLeuSerSerAla 27

```
seq_documentation_block:
  ID: O99H82
  PRELIMINARY:
    PRT: 1213 AA
```

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE GENOMIC DNA, CHROMOSOME

OS *Arabidopsis thaliana* (Mouse-ear cress).

OC spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; fabales; fabaceae; papilionaceae; viciae

CC eucosids II; Brassicales; Brassicaceae; Arabidopsis

OX . NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.,

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RX PubMed=10907853;

RA Nakamura Y.;

RT "Structural analysis"

RT Sequence features

RT TAC and BAC clones:

RL DNA Res. 7:217-221(2000).
 DR EMBL; AF002044; BAB02250.1; ..
 DR InterPro; IPR000886; ER_target.
 DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
 SQ SEQUENCE 1213 AA; 133944 MW; 1EAC2E38474EFD01 CRC64;

alignment_scores:
 Quality: 57.00 Length: 12
 Ratio: 5.182 Gaps: 0
 Percent Similarity: 91.667 Percent Identity: 75.000

alignment_block:

US-09-049-696-17 x Q9LHK2 ..

Align seg 1/1 to: Q9LHK2 from: 1 to: 1213

25 TGGAGTGGATAGGAGAACTGCAGCTGTCAATAGCC 60
 |||::|||::|||::|||::|||::|||::|||::|||
 296 TrpArgTrpValGlyGluLeuGlnLeuSerSerAla 307

seq_name: sp_invertebrate:Q9NJG3

seq_documentation_block:

ID Q9NJG3 PRELIMINARY; PRT; 27 AA.

AC Q9NJG3;

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE LABIAL (FRAGMENT).

GN LAB.

OS Steatoda triangulosa.

OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;

OC Araneomorphae; Entelegynae; Araneioidea; Theridiidae; Steatoda.

OX NCBI_TaxID=114396;

RN [1]

RP SEQUENCE FROM N.A.

RA Abzhanov A., Popadic A., Kaufman T.C.;

RT "Chelicerate Hox genes and the homology of arthropod segments.";

RL Evol. Dev. 0:0-0(2000).

DR EMBL; AF151987; AAF73203.1; ..

DR InterPro; IPR001356; Homeobox.

DR Pfam; PF00046; homeobox; 1.

FT NON_TER 1

FT NON_TER 27

SQ SEQUENCE 27 AA; 3159 MW; 888ED462476DA25D CRC64;

alignment_scores:

Quality: 52.00 Length: 25
 Ratio: 3.059 Gaps: 0
 Percent Similarity: 68.000 Percent Identity: 44.000

alignment_block:

US-09-049-696-17 x Q9NJG3 ..

Align seg 1/1 to: Q9NJG3 from: 1 to: 27

9 CATTTTAAATAATTATGTGAAGTGGATAGGAGAACTGCAGCTGTCAATAG 58
 |||::|||::|||::|||::|||::|||::|||::|||
 1 HisPheAsnLysTrpLeuThrArgAlaAArgGileGluileAlaSer 17

59 CCTAGGCTGAATTTTGTGCAGATA 83

|||||::|||::|||::|||::|||::|||::|||::|||

17 rLeuGlyLeuAsnGluThrGlnVal 25

seq_name: sp_bacteria:Q9XAU2

seq_documentation_block:

ID Q9XAU2 PRELIMINARY; PRT; 199 AA.

AC Q9XAU2;

DT 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOTHETICAL 21.5 KDA PROTEIN.
 OS Pseudomonas stutzeri (Pseudomonas perfectomarina).
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=316;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JM300;
 RA Graupner S., Wackernagel W.;
 RT "unpublished";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DDJB databases.
 DR EMBL; AJ243354; CAB50918.1; ..
 DR InterPro; IPR002898; MOTA_ExbB.
 DR Pfam; PF01618; MOTA_ExbB; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 199 AA; 21482 MW; 6371F17B93DEAB3 CRC64;

alignment_scores:

Quality: 52.00 Length: 16
 Ratio: 4.333 Gaps: 0
 Percent Similarity: 75.000 Percent Identity: 56.250

alignment_block:

US-09-049-696-17 x Q9XAU2 ..

Align seg 1/1 to: Q9XAU2 from: 1 to: 199

7 CACATTTTAAATAATTATGTGAAGTGGATAGGAGAACTGCAGCTGTCA 54
 |||::|||::|||::|||::|||::|||::|||::|||
 31 HisLeuLeuGlyGlnValTrpLysTrpLysAspLysGlnLeuSer 46

seq_name: sp_rodent:Q9WU91

seq_documentation_block:

ID Q9WU91 PRELIMINARY; PRT; 342 AA.

AC Q9WU91;

DT 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE ODOBRANT RECEPTOR S25.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BALB/C;

RX MEDLINE-99189756; PubMed-10089886;

RA Malnic B., Hirono J., Sato T., Buck L.B.;

RT "Gubernatorial receptor codes for odors.";

RL Cell 96:713-723(1999).

DR EMBL; AF121977; AAD27597.1; ..

DR InterPro; IPR000276; GPCR_Rhodpsn.

DR Pfam; PF00001; 7tm_1; 1.

DR PRINTS; PR00237; GPCRHHODOPSN.

DR PROSITE; PS00237; G_PROTEIN_RECF_F1_1; UNKNOWN_1.

DR PROSITE; PS00262; G_PROTEIN_RECF_F1_2; 1.

KW Receptor.

SQ SEQUENCE 342 AA; 38265 MW; 96F32E89C275D04F CRC64;

alignment_scores:

Quality: 52.00 Length: 14
 Ratio: 4.333 Gaps: 0
 Percent Similarity: 85.714 Percent Identity: 71.429

alignment_block:

US-09-049-696-17 x Q9WU91 ..

Align seg 1/1 to: Q9WU91 from: 1 to: 342

Percent Similarity: 73.077 Percent Identity: 50.000

alignment_block:

US-09-049-696-17/rev x Q9BI48 ..

Align seg 1/1 to: Q9BI48 from: 1 to: 3734

93 TTATTTTATTATCTGACAAAAT...TCAGCCCTAGGCTATTGACAGCT 47
 :::::::::::::: ||| :::::||| ||| ||||
 2601 MetPheTyrLeuSerGlnMetAsnGluAsnSerLeuGluTyr...GlnLe 2616

46 GCAGTTCTCCTATCCACTCCACATAAT 19
 ||| |||:::|||||:::
 2616 uProPheIleGlnProPheProHisSer 2625

This Page Blank (uspto)

OM of: US-09-049-696-17 to: SwissProt_39:* out_format : pfs
Date: Mar 30, 2002 2:52 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODE=frame+nt2p.model -DEV=xlp
-Q/cgnt_1/USPTO.spool/US09049696/runat_28032002_145238_2085/app_query.fasta_1.12579
-DB=SwissProt_39 -QFMT=FASTAN -SUFFIX=isp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsmb2
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=EXT -MINLEN=0 -MAXLEN=2000000000
-USER=US09049696 @CGN1_1_165 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1.

Search information block:
Query: US-09-049-696-17
Query length: 106
Database: SwissProt_39:*
Database sequences: 100059
Database length: 36664827
Search time (sec): 306.030000

score_list:	Sequence	Strd Orig	ZScore	EScore	Len	! Documentation
	SwissProt_39:SOX5_HUMAN	+	56.00	143.00	0.7832	347
	SwissProt_39:RHA1_ECOLI	+	52.00	129.93	3.47	419
	SwissProt_39:HX1A_MAIZE	+	52.00	125.63	3.51	719
	SwissProt_39:YD9A_METJA	+	52.00	123.12	3.53	987
	SwissProt_39:VGLB_HSVJA	+	50.50	120.37	6.13	808
	SwissProt_39:IAP1_NPVAC	+	50.00	127.17	7.24	286
	SwissProt_39:HXD1_XENJA	+	49.00	123.90	10.50	300
	SwissProt_39:SOX5_MOUSE	+	49.00	121.77	10.55	332
	SwissProt_39:SSK2_YEAST	+	49.00	110.70	10.84	1579
	SwissProt_39:SP1A_VACCC	+	48.00	128.56	14.93	116
	SwissProt_39:VBR1_PHUV	+	48.00	122.27	15.17	256
	SwissProt_39:YH86_SNY3	+	48.00	122.11	15.17	261
	SwissProt_39:RG55_DROME	+	47.50	119.61	18.31	298
	SwissProt_39:HMB3_TRIGR	+	47.00	126.23	21.62	108
	SwissProt_39:KAPB_PIG	+	47.00	117.16	22.10	338
	SwissProt_39:KAPB_BOVIN	+	47.00	116.89	22.12	350
	SwissProt_39:PAR3_HUMAN	+	47.00	116.36	22.15	374
	SwissProt_39:KAP1_BOVIN	+	47.00	115.89	22.17	397
	SwissProt_39:AAP1_YEAST	+	47.00	109.78	22.51	856
	SwissProt_39:OL7E_MOUSE	+	46.00	123.05	31.35	112
	SwissProt_39:HME1_BRARE	+	46.00	117.30	31.80	231
	SwissProt_39:CDX2_MOUSE	+	46.00	114.93	31.99	311
	SwissProt_39:PAR3_MOUSE	+	46.00	113.57	32.09	369
	SwissProt_39:YEA1_ECOLI	+	46.00	111.30	32.27	491
	SwissProt_39:HMLA_DROME	+	46.00	109.26	32.43	635
	SwissProt_39:SLI7_ENTHI	+	46.00	103.65	32.88	1285
	SwissProt_39:IF2C_GUTH	+	45.50	106.65	39.16	735
	SwissProt_39:BGAL_ASPOF	+	45.50	105.66	39.26	832
	SwissProt_39:ZNUC_BUCAI	+	45.00	114.17	46.12	238
	SwissProt_39:Y41H_RHISN	+	45.00	114.13	46.12	239
	SwissProt_39:HME3_BRARE	+	45.00	113.43	46.20	261
	SwissProt_39:YF85_METJA	+	45.00	112.06	46.36	310
	SwissProt_39:O3A2_HUMAN	+	45.00	111.94	46.37	315
	SwissProt_39:O3A3_HUMAN	+	45.00	111.94	46.37	315
	SwissProt_39:CKR2_HUMAN	+	45.00	110.57	46.53	374
	SwissProt_39:GNCR_MOUSE	+	45.00	110.28	46.56	388
	SwissProt_39:COX8_HUMAN	+	45.00	109.23	46.68	443
	SwissProt_39:SPN5_SCHPO	+	45.00	108.86	46.73	464
	SwissProt_39:6PGD_ACTAC	+	45.00	108.52	46.76	484
	SwissProt_39:CFLA_MOUSE	+	45.00	108.52	46.76	484

SwissProt_39:VN53_ROTBR + 45.00 108.41 46.78 491 ! P12475 bovine rotavirus (

SwissProt_39:VG29_BPMU + 45.00 108.08 46.82 512 ! O91W5 bacteriophage mu. p

SwissProt_39:GIC4_SOYBN + 45.00 107.33 46.90 562 ! P02858 glycine max (soybe

SwissProt_39:SPB4_YEAST + 45.00 106.74 46.97 606 ! P25808 saccharomyces cere

SwissProt_39:SPEL_ARATH + 45.00 105.57 47.10 702 ! Q9S164 arabidopsis thalia

seq_name: SwissProt_39:SOX5_HUMAN

seq_documentation_block:

ID SOX5_HUMAN STANDARD; PRT; 347 AA.

AC P35711;

DC 01-JUN-1994 (Rel. 29, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE 20-AUG-2001 (Rel. 40, Last annotation update)

DE TRANSCRIPTION FACTOR SOX-5.

GN SOX5.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RX MEDLINE=96411696; PubMed=8812465;

RA Wunderle V.M., Critcher R., Ashworth A., Goodfellow P.N.;

RT "Cloning and characterization of SOX5, a new member of the human SOX

RT gene family.";

RL Genomics 36:354-358(1996).

RN [2]

RP SEQUENCE OF 151-204 FROM N.A.

RX MEDLINE=92310993; PubMed=1614875;

RA Denny P., Swift S., Brand N., Dabhadre N., Barton P., Ashworth A.;

RT "A conserved family of genes related to the testis determining gene,

RT SRV.";

RL Nucleic Acids Res. 20:2887-2897(1992).

CC -1- FUNCTION: BINDS SPECIFICALLY TO THE DNA SEQUENCE 5'-AACAAAT-3'.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).

CC -1- TISSUE SPECIFICITY: FOUND IN ALL TISSUES TESTED.

CC -1- SIMILARITY: CONTAINS 1 HMG BOX.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; S83308; AAB49537.1; -

DR EMBL; X65658; CAA46609.1; -

DR EMBL; X65662; CAA46613.1; -

DR PIR; S21480; S21480.

DR PIR; S22939; S22939.

DR PIR; S21485; S21485.

DR HSSP; Q05066; 1HRZ.

DR TRANSFAC; T01861; -.

DR MIM; 604975; -.

DR InterPro; IPR000910; HMG_12_box.

DR Pfam; PF00505; HMG_box; 1.

DR SMART; SM00398; HMG; 1.

DR DNA-binding; Nuclear protein; Transcription regulation.

KW DNA_BIND 140 208 HMG BOX.

SQ SEQUENCE 347 AA; 38848 MW; 16D36F5D8B105A2E CRC64;

alignment_scores:

Quality: 56.00 Length: 20

Ratio: 3.294 Gaps: 0

Percent Similarity: 85.000 Percent Identity: 55.000

alignment_block:

US-09-049-696-17 x SOX5_HUMAN

Align seg 1/1 to: SOX5_HUMAN from: 1 to: 347

```
30 GTGATAGGAGAACTGCTGCTCAATGCTAGGCTGCTCAATTTCTCA 79
   :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 LeuAspGlyValAlaValAlaValAsnSerLeuGlyLeuAsnGlyCysAr 77
   :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
80 CATAAATAAA 89
   :   :|||
77 gThrGluLys 80
```

seg_name: SwissProt_39:RHAA_ECOLI

seg_documentation_block:

```
ID RHAA_ECOLI STANDARD; PRT; 419 AA.
AC P32170;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE L-RHAMNOSE ISOMERASE (EC 5.3.1.14).
GN RHAA OR B3903.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
   SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=93347969; PubMed=8346018;
RA Plunkett G., III, Burland V.D., Daniels D.L., Blattner F.R.;
RT "Analysis of the Escherichia coli genome. III. DNA sequence of the
   region from 87.2 to 89.2 minutes."
RL Nucleic Acids Res. 21:3391-3398(1993).
RN [2]
   SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93374854; PubMed=8396120;
RA Moralejo P., Egan S.M., Hidalgo E.F., Aguilar J.;
RT "Sequencing and characterization of a gene cluster encoding the
   enzymes for L-thiamine metabolism in Escherichia coli."
RL J. Bacteriol. 175:5585-5594(1993).
RN [3]
   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX MEDLINE=20351473; PubMed=10891278;
RA Korndorfer I.P., Fessner W.D., Matthews B.W.;
RT "The structure of rhamnose isomerase from Escherichia coli and its
   relation with xylose isomerase illustrates a change between inter and
   intra-subunit complementation during evolution."
RL J. Mol. Biol. 300:917-933(2000).
CC -1- CATALYTIC ACTIVITY: L-RHAMNOSE = L-RHAMNULOSE.
CC -1- COFACTOR: MANGANESE.
CC -1- PATHWAY: FIRST STEP IN THE METABOLISM OF L-RHAMNOSE.
CC -1- SUBUNIT: HOMOTETRAMER.
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

```
EMBL; L19201; AAB03036.1; ..
DR EMBL; AE000465; AAC76885.1; ..
DR EMBL; X60472; CAM3002.1; ..
DR PIR; S40847; S40847.
DR PDB; IDE6; 09-AUG-00.
DR PDB; IDE5; 09-AUG-00.
DR EcoGene; EG11867; rhaA.
KW Rhamnose metabolism; Isomerase; Manganese; 3D-structure;
KW Complete proteome.
FT CONFLICT 359 360 DV -> EL (IN REF. 2).
FT CONFLICT 366 366 A -> P (IN REF. 2).
FT SEQUENCE 419 AA; 47171 MW; D387E941C94883A1 CRC64;
```

alignment_scores:

```
Quality: 52.00 Length: 22
Ratio: 3.250 Gaps: 1
Percent Similarity: 72.727 Percent Identity: 54.545
```

alignment_block:

US-09-049-696-17 x RHAA_ECOLI ..

Align seg 1/1 to: RHAA_ECOLI from: 1 to: 419

```
9 CATTTTAAAAATTATGCTGAAGTGGATAGGAGAACTGCAGCTGCTCAATAG 58
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
114 HisPheLysAsnTrpValGlu.....TrpAlaLysAlaAsnG1 126
```

59 CCTAGGCTGAATTT 74

:|||||:|||||:|||||

126 nLeuGlyLeuAspPhe 131

seg_name: SwissProt_39:HX1A_MAIZE

seg_documentation_block:

ID HX1A_MAIZE STANDARD; PRT; 719 AA.

AC P46605;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE HOMEBOX PROTEIN HOX1A.

GN HOX1A.

OS Zea mays (Maize).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;

OC Panicoidae; Andropogoneae; Zea.

OX NCBI_TaxID=4577;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Shoot;

RX MEDLINE=92371444; PubMed=1354614;

RA Bellmann R., Werr W.;

RT "Zmhox1a, the product of a novel maize homeobox gene, interacts with

the shrunken 26 bp feedback control element.";

RL EMBO J. 11:3367-3374(1992).

CC -1- FUNCTION: INTERACTS WITH THE SHRUNKEN (SHR) 26 BP FEEDBACK CONTROL

ELEMENT.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN KERNELS, LEAVES AND SHOOTS BUT

NOT IN ROOTS.

CC -1- SIMILARITY: BELONGS TO THE PHD-ASSOCIATED FAMILY OF HOMEBOX

PROTEINS.

CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; X67561; CAA47859.1; ..

DR HSSP; P22808; 1VND.

DR TRANSFAC; T00922; ..

DR MaizeDB; 25707; ..

DR InterPro; IPR000047; HTH_repressr.

DR InterPro; IPR001356; Homeobox.

DR InterPro; IPR001965; PHD.

DR InterPro; IPR001841; Znf_ring.

DR Pfam; PF00046; homeobox; 1.

DR Pfam; PF00628; PHD; 1.

DR PRINTS; PR00031; HTHREPRESSR.

DR SMART; SM00389; HOX; 1.

DR SMART; SM00249; PHD; 1.

SQ SEQUENCE 719 AA; 79116 MW; 5536608CA71CB318 CRC64;

alignment_block:

591 rLeuAlaGluGluLeuGlyLeuThrPheArgGlnValAsnLys 605

100

```

FT CARBOHYD 158 158 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 355 355 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 569 569 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 598 598 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 808 AA; 91694 MW; FA1BCB89B9D5672E CRC64;

```

```

alignment_scores:
  Quality: 50.50 Length: 36
  Ratio: 2.196 Gaps: 2
  Percent Similarity: 63.889 Percent Identity: 41.667

```

alignment_block:

US-09-049-696-17 x VGLB_HSVSA ..

Align seg 1/1 to: VGLB_HSVSA from: 1 to: 808

```

9 CATTGTTAAATTTATGCGAAGTGGATAGGAGAACTGCAGCTGTCAATAG 58
|||||
573 HisPheLysAsnTyrValHisValGlu.....ThrLeuProValAsnAs 587
59 C.....CTAGGGCTGAATTTTGTGCAGATAATA 87
|||||
587 nileSerThrLeuAspThrPheLeuAlaLeuAsnLeuThrPheIleGluA 604
88 AAATAAAT 95
|||||
604 snIleAsp 606

```

seq_name: SwissProt_39:IAP1_NPVAC

```

seq_documentation_block:
ID IAP1_NPVAC STANDARD; PRT; 286 AA.
AC P41435;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE APOPTOSIS INHIBITOR 1 (IAP-1).
GN IAP1.
OS Autographa californica nuclear polyhedrosis virus (AcMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=46015;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C6;
RX MEDLINE=94303173; PubMed=8030224;
RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RT "The complete DNA sequence of Autographa californica nuclear
polyhedrosis virus."
RL Virology 202:586-605(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=E2;
RX MEDLINE=93079853; PubMed=1333113;
RA Braunagel S.C., Daniel K.D., Reilly L.M., Guarino L.A., Hong T.,
RA Summers M.D.;
RT "Sequence, genomic organization of the EcoRI-A fragment of Autographa
californica nuclear polyhedrosis virus, and identification of a
viral-encoded protein resembling the outer capsid protein VP8 of
rotavirus."
RL Virology 191:1003-1008(1992).

```

```

CC -!- FUNCTION: ACTS BY BLOCKING CELLULAR APOPTOSIS RATHER THAN BY
PREVENTING VIRAL STIMULATION OF APOPTOSIS.
CC -!- SIMILARITY: CONTAINS 2 BIR REPEATS.

```

```

CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

```

DR EMBL; L22858; AAA66657.1; -.
DR EMBL; M96361; AAA66796.1; -.
DR PIR; D36828; D36828.
DR InterPro; IPR001370; BIR.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00653; BIR; 2.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00238; BIR; 2.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS01282; BIR_REPEAT_1; 2.
DR PROSITE; PS0143; BIR_REPEAT_2; 2.
DR APOPTOSIS; Zinc-finger; Repeat.
FT REPEAT 29 96 BIR 1.
FT REPEAT 131 199 BIR 2.
FT ZN_FING 238 273 RING-TYPE.
SQ SEQUENCE 286 AA; 33320 MW; FFEE505A35EF1BEA CRC64;

```

```

alignment_scores:
  Quality: 50.00 Length: 32
  Ratio: 2.273 Gaps: 2
  Percent Similarity: 68.750 Percent Identity: 40.625

```

alignment_block:

US-09-049-696-17 x IAP1_NPVAC ..

Align seg 1/1 to: IAP1_NPVAC from: 1 to: 286

```

3. CATTCACTTTTAAATTTAT...GTGGAAGTGGATAGGAGAACTGCAGC 49
|||||
33 HisSerPheGluAsnTyrProIleGlu.....AsnThrAlaPh 46
50 TGTCATAGCTAGGCTGGAATTTTGTGCAGATAATAATAAAT 95
|||||
46 eileAsnSerLeuIleValAsnGlyPheLysTyrAsnGlnValasp 61

```

seq_name: SwissProt_39:HXDL_XENLA

```

seq_documentation_block:
ID HXDL_XENLA STANDARD; PRT; 300 AA.
AC Q08820; Q91772;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HOMEOBOX PROTEIN HOX-D1 (HOX.LAB1).
GN HCXDL.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Koim P.J., Frey J., Sive H.L.;
RA Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 75-300 FROM N.A.
RX MEDLINE=91331304; PubMed=1678362;
RA Sive H.L., Cheng P.F.;
RT "Retinoic acid perturbs the expression of xhox.lab genes and alters
mesodermal determination in Xenopus laevis."
RL Genes Dev. 5:1321-1332(1991).

```

```

CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF

```

```
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC ACTS ON THE ANTERIOR BODY STRUCTURES.
CC -|- SUBCELLULAR LOCATION: NUCLEAR.
CC -|- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.
CC "LABIAL" SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L25857; AAA03480.1; -.
CC EMBL; L25857; AAA03480.1; -.
CC HSSP; P02833; 1SAN.
CC InterPro; IPR001827; Antennapedia.
CC InterPro; IPR001356; Homeobox.
CC Pfam; PF00046; homeobox; 1.
CC PRINTS; PR00024; HOMEBOX.
CC PRINTS; PR00025; ANTENNAPEIDIA.
CC SMART; SM00389; Hox; 1.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS00071; HOMEBOX_2; 1.
CC Homeobox; DNA-binding; Developmental protein; Nuclear protein;
CC Transcription regulation.
CC DOMAIN 175 180 ANTP-TYPE HEXAPEPTIDE.
CC FT DNA_BIND 200 259 HOMEBOX.
CC FT CONFLICT 75 75 P -> A (IN REF. 2).
CC FT CONFLICT 291 300 TLSPKQASP -> TRRLGRFTLETOSKPNMWDCLLHPK
CC FT CONFLICT 291 300 OPKL (IN REF. 2).
CC SEQUENCE 300 AA; 33343 MW; A1565BCB55AF1810 CRC64;

alignment_scores:
  Quality: 49.00 Length: 25
  Ratio: 3.062 Gaps: 0
Percent Similarity: 64.000 Percent Identity: 44.000

alignment_block:
US-09-049-696-17 x HXD1_XENLA ..
Align seg 1/1 to: HXD1_XENLA from: 1 to: 300
9 CATTTTAAATTTATGTGGAGCTGGATGAGGAAGTGCAGCTGTCGAATAG 58
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 HisPheAsnLysTyrLeuThrArgAlaArgIleGluIleAlaAsnSe 236

59 CCTAGGCGTGAATTTTGTGCAGATA 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 rLeuGlnLeuAsnAspThrGlnVal 244

seq_name: SwissProt_39:SOX5_MOUSE
seq_documentation_block:
ID SOX5_MOUSE STANDARD; PRT; 392 AA.
AC P35710;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TRANSCRIPTION FACTOR SOX-5.
GN SOX5 OR SOX-5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=93010962; PubMed=1396566;
RA Denny P., Swift S., Connor F., Ashworth A.;
RT "An SRY-related gene expressed during spermatogenesis in the mouse
```

```
RT encodes a sequence-specific DNA-binding protein.";
RL EMBL J. 11:3705-3712(1992).
RN [2]
EX CHARACTERIZATION.
RA Cqnor F., O'Carry P.D., Read C.M., Preston N.S., Driscoll P.C.,
RA Denny P., Crane-Robinson C., Ashworth A.;
RA DNA binding and bending properties of the post-meiotically expressed
RA SRY-related protein Sox-5.";
RL Nucleic Acids Res. 22:3339-3346(1994).
CC -|- FUNCTION: BINDS SPECIFICALLY TO THE DNA SEQUENCE 5'-AACAAAT-3'.
CC -|- SUBCELLULAR LOCATION: NUCLEAR.
CC -|- TISSUE SPECIFICITY: TESTIS.
CC -|- DEVELOPMENTAL STAGE: EXPRESSED DURING SPERMATOGENESIS.
CC -|- SIMILARITY: CONTAINS 1 HMG BOX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X65657; CAA46608.1; -.
CC EMBL; X65657; CAA46608.1; -.
CC HSSP; Q05066; 1HRZ.
CC TRANSFAC; T01429; -.
CC MGD; MGI:98367; Sox5.
CC InterPro; IPR000910; HMG_12_box.
CC Pfam; PF00505; HMG_box; 1.
CC SMART; SM00398; HMG; 1.
CC DNA-binding; Nuclear protein; Transcription regulation.
CC DOMAIN 185 253 POLY-ALA.
CC FT DNA_BIND 370 373 POLY-GLU.
CC FT DOMAIN 392 AA; 43203 MW; 0B7F33981A390C12 CRC64;
CC SEQUENCE 392 AA; 43203 MW; 0B7F33981A390C12 CRC64;

alignment_scores:
  Quality: 49.00 Length: 20
  Ratio: 2.882 Gaps: 0
Percent Similarity: 85.000 Percent Identity: 45.000

alignment_block:
US-09-049-696-17 x SOX5_MOUSE ..
Align seg 1/1 to: SOX5_MOUSE from: 1 to: 392
30 GTGGATAGGAGAACTGCACCTGTCATACCTAGGCTGCAATTTTGTCA 79
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
106 LeuAspGlyLysValAlaValValAsnSerIleGlyLeuSerAsnCysAr 122

80 GATAAATAAA 89
: : : : :
122 gThrGluLys 125

seq_name: SwissProt_39:SSK2_YEAST
seq_documentation_block:
ID SSK2_YEAST STANDARD; PRT; 1579 AA.
AC P35599;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAP KINASE KINASE KINASE SSK2 (EC 2.7.1.-) (SUPPRESSOR OF SENSOR
DE KINASE 2).
GN SSK2 OR YNR031C OR N3276.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
```

```

DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE SERINE PROTEINASE INHIBITOR 2 HOMOLOG FIRST PART.
GN B13R.
OS Vaccinia virus (strain Copenhagen).
OC Vfususes: dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OC NCBI_TaxID=10249;
[1]
RN SEQUENCE FROM N.A.
RN MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Winslow J.P.,
RA Paoletti E.;
RA "the complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
[2]
RN COMPLETE GENOME.
RN Goebel S.J., Johnson G.P., Perkus M.E., Winslow J.P.,
RA Paoletti E.;
RA Appendix to 'the complete DNA sequence of vaccinia virus'.;
RL Virology 179:517-563(1990).
CC -!- FUNCTION: THIS PROTEIN IS NON-FUNCTIONAL. IN CONTRAST TO STRAIN
CC WR, WHERE THE SPI-2 INHIBITOR MAY BE INVOLVED IN THE REGULATION
CC OF THE COMPLEMENT CASCADE. IN STRAIN COPENHAGEN IT IS INTERRUPTED
CC BY A STOP CODON AND THUS IS FOUND IN TWO PARTS.
CC -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or: send an email to license@isb-sib.ch).
-----
DR EMBL; M35027; AAA48210.1; -.
DR F1R; C42527; WMVZB3.
DR HSSP; P05619; LHLE.
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin; 1.
DR Serpin; Hypothetical protein.
KW SEQUENCE 116 AA; 12849 MW; 8A2E888330749B74 CRC64;
SQ
alignment_scores:
      Quality: 48.00      Length: 24
      Ratio: 2.400      Gaps: 0
      Percent Similarity: 83.333      Percent Identity: 37.500
alignment_block:
US-09-049-696-17 x SPIA_VACCC ..
Align seq 1/1 to: SPIA_VACCC from: 1 to: 116
      21 TATGTGGAAGTGGATAGGAACTGCAGCTGCTCAATAGCTAGGCTGAA 70
      : ||||| ||||| : ||||| : ||||| : ||||| : ||||| :
      35 TyrValGluThrGluAsnThrAspLysValSerAlaGlnAsnIleSe 51
      : ||| : ||||| : ||||| : ||||| : ||||| : ||||| :
      71 TTTTGTGCAGATAAATAAATA 92
      : ||| : ||||| : ||||| : ||||| : ||||| : ||||| :
      51 rPheLysSerIleAsnLysVal 58
seq_name: SwissProt_39:VBR1_PHVU
seq_documentation_block:
ID VBR1_PHVU STANDARD; PRT; 256 AA.
AC Q06927;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE BRL1-PROTEIN.
DE Pepper huasteco virus (PHV).
GN BRL1.

```


OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=28349;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94015007; PubMed=8409944;
RA Torres-Pacheco I., Garzon-tiznado J.A., Herrera-Estrella L.,
RA Rivera-Bustamante R.F.;
RT "Complete nucleotide sequence of pepper huasteco virus: analysis and
RT comparison with bipartite geminiviruses.";
RL J. Gen. Virol. 74:2225-2231(1993).

CC -!- SIMILARITY: BELONGS TO GEMINIVIRUSES BR1 PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; X70419; CAA49860.1;
DR PIR; S31872; S31872.
DR PIR; JQ2304; JQ2304.
DR InterPro; IPR001530; Gemini_BR1.
DR InterPro; IPR000263; Gemini_coat_AR1_BR1.
DR Pfam; PF01489; Gemini_BR1; 1.
DR PRINTS; PR00223; GEMCOATBR1.
DR PRINTS; PR00225; GEMCOATBR1.
SQ SEQUENCE 256 AA; 29529 MW; B6B0FCF684612730 CRC64;

alignment_scores:
Quality: 48.00 Length: 21
Ratio: 3.200 Gaps: 0
Percent Similarity: 71.429 Percent Identity: 42.857

alignment_block:
US-09-049-696-17 x VBRL_PHUV ..

Align seg 1/1 to: VBRL_PHUV from: 1 to: 256

9 CATTTAAATAATGTGTGAAGTGGATGAGGAAGTGCAGCTGTCAATAG 58
|||||
100 HisPheIysGlyThrLeuIysIleGluArgValGlySerGluValAsnMe 116
59 CCTAGGCGCTCAAT 71
|||||
116 talaglyLeuAsn 120

seq_name: SwissProt_39:YH86_SYNY3

seq_documentation_block:
ID YH86_SYNY3 STANDARD; PRT; 261 AA.
AC P73335;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE DEOXYRIBONUCLEASE SLI1786 (EC 3.1.21.-).
GN SLI1786.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";

RL DNA Res. 3:109-136(1996).
CC -!- SIMILARITY: BELONGS TO THE TATD DNASE FAMILY.

CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; D90905; BAA17366.1;
DR InterPro; IPR001130; UPF0006.
DR Pfam; PF01026; TATD_DNase; 1.
DR PROSITE; PS01137; TATD_1; 1.
DR PROSITE; PS01090; TATD_2; FALSE_NEG.
DR PROSITE; PS01091; TATD_3; 1.
KW Hypothetical protein; Hydrolase; Nuclease; Complete proteome.
SQ SEQUENCE 261 AA; 29258 MW; 7F897DF2FE5131A2 CRC64;

alignment_scores:
Quality: 48.00 Length: 26
Ratio: 2.182 Gaps: 0
Percent Similarity: 84.615 Percent Identity: 30.769

alignment_block:
US-09-049-696-17 x YH86_SYNY3 ..

Align seg 1/1 to: YH86_SYNY3 from: 1 to: 261

21 TATGTGAGTGGATGAGGAAGTGCAGCTGTCAATAGCTAGGCTGAA 70
|||||
82 TyrAlaIysAlaAspArgValAlaIleGlyGluMetGlyLeuAs 98
71 TTTTGTGCAGATAATAATAAATCAT 98
|||||
98 pPhePheLysAlaAspAsnArgAspHis 107

seq_name: SwissProt_39:REG5_DROME

seq_documentation_block:
ID REG5_DROME STANDARD; PRT; 298 AA.
AC Q94913;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE RHYTHMICALLY EXPRESSED GENE 5 PROTEIN (DREG-5).
GN REG-5.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S; TISSUE=Head;
RX MEDLINE=96203080; PubMed=8612586;
RA van Gelder R.N., Krasnow M.A.;
RT "A novel circadianly expressed Drosophila melanogaster gene dependent
RT on the period gene for its rhythmic expression.";
RL EMBO J. 15:1625-1631(1996).
CC -!- FUNCTION: INVOLVED IN THE GENERATION OF BIOLOGICAL RHYTHMS
CC (POTENTIAL). IN THE HEAD, OSCILLATES IN ABUNDANCE WITH A DAILY
CC PEAK DURING EARLY NIGHT, EVEN UNDER CONSTANT DARKNESS. OSCILLATION
CC IS DEPENDENT ON PERIOD (PER) FUNCTION.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN HEAD, BUT NOT IN THE BODY.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN 24 HOURS EMBRYO.

CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

DR PROSITE: PS00108; PROTEIN_KINASE_ST: 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transferase; Serine/threonine-protein kinase; ATP-binding; CAMP;
 KW Phosphorylation.
 FT NON_TER 1 1
 FT DOMAIN 31 285 PROTEIN KINASE.
 FT NP_BIND 37 45 ATP (BY SIMILARITY).
 FT BINDING 60 60 ATP (BY SIMILARITY).
 FT ACT_SITE 154 154 BY SIMILARITY.
 FT MOD_RES 185 185 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 326 326 PHOSPHORYLATION (BY SIMILARITY).
 SQ SEQUENCE 338 AA; 39348 MW; 7BD356357FD3BEFE CRC64;

alignment_scores:
 Quality: 47.00 Length: 25
 Ratio: 2.611 Gaps: 0
 Percent Similarity: 72.000 Percent Identity: 36.000

alignment_block:
 US-09-049-696-17 x KAPB_PIG ..
 Align seg 1/1 to: KAPB_PIG from: 1 to: 338
 6 TCACATTTTAAATATGTGGAGTGGATAGGAGACTCCAGCTGTCAA 55
 |||:|||| :||| ||| |||:|||| :|||
 313 SerAsnPhcAspTyrGluGluAspIleArgValSerIleThrGI 329
 56 TAGCCTAGGGCTGAATTTTGTTCAG 80
 :||| :|||:|||||:
 329 uLysCysGlyLysGluPheCysGlu 337

This Page Blank (uspto)

OM of: US-09-049-696-17 to: PIR_68.* out_format : pfs

Date: Mar 30, 2002 2:27 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+ntp.model -DEV=slp
-Q/cn2_1/USPTO.spool/US09049696/runat_28032002_145238_2031/app_query.fasta_1.12579
-DB=PIR_68 -GFM=fastan -SUFFIX=pr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.col
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09049696_@CGN1_1_367 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-049-696-17
Query length: 106
Database: PIR_68.*
Database sequences: 219241
Database length: 76174552
Search time (sec): 470.790000
```

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
pir2:JG0168	+	90.00	233.41	913	gob-5 protein - mouse
pir2:S40847	+	52.00	129.68	419	L-rhamnose isomerase (EC 5.3.1.1)
pir2:EB6079	+	52.00	129.68	419	L-rhamnose isomerase [imported]
pir2:S25237	+	52.00	125.34	719	homeotic protein Hox1a - maize
pir2:B84741	+	52.00	125.34	719	hypothetical protein At2g33080
pir2:A64474	+	52.00	125.11	740	hypothetical protein MJ1394 - M
pir2:EB4463	-	51.50	125.13	987	mutator-like transposase [imported]
pir2:VBGESM	+	50.50	120.06	616	glycoprotein B precursor - salmon
pir2:T30149	+	50.00	126.96	808	orf13 protein - Autographa cal
pir2:T03377	+	50.00	124.45	286	hypothetical protein C39H7.4 -
pir2:B81711	-	50.00	119.86	391	homeotic protein Hox1B - maize
pir2:T12983	-	49.50	120.14	692	conserved hypothetical protein
pir2:A40357	+	49.50	117.88	558	hypothetical protein T21f8.30 -
pir2:T41772	+	49.00	124.47	739	homeotic protein lab1 - African
pir2:C86452	+	49.00	123.90	272	IAP1 orf27 - Bombyx mori nuclea
pir2:T27712	+	49.00	122.99	327	protein F6N18.11 [imported] - A
pir2:S25195	+	49.00	123.35	354	hypothetical protein ZK1225.1 -
pir2:T9179	+	49.00	121.53	392	sex-determining protein SOX-5 -
pir2:S59801	-	49.00	115.35	846	hypothetical protein C10C5.1 -
pir2:T34238	+	49.00	110.33	1579	protein kinase SSK2 (EC 2.7.1.1)
pir2:D82569	-	48.00	126.97	139	hypothetical protein F2A3.6 -
pir2:C85626	+	48.00	125.59	165	conserved hypothetical protein
pir2:J02304	+	48.00	122.31	248	hypothetical protein Z1338 [imp
pir2:S31872	+	48.00	122.06	256	BL1 protein - pepper huasteco v
pir2:S77519	+	48.00	121.90	256	BL1 protein - pepper rizado ama
pir2:H2160	+	48.00	117.32	261	hypothetical protein s111786 -
pir2:T51434	+	48.00	113.61	462	conserved hypothetical protein
pir2:S72278	+	48.00	113.42	733	hypothetical protein F2G14.10 -
pir2:I52300	+	48.00	113.26	765	ATP-dependent Clp proteinase (B
pir2:A56539	+	48.00	101.70	3225	giantin - human
pir2:S9241	+	48.00	101.62	3259	giantin - human
pir2:S61505	+	47.50	119.39	3259	Dreg-5 protein - fruit fly (Dre
pir2:S05258	+	47.50	111.95	298	UDPGlucose--starch glucosyltran
pir2:S00085	+	47.00	126.10	752	homeotic protein HB3 - sea urch
pir2:ORX081	+	47.00	116.93	108	protein kinase (EC 2.7.1.37), c
pir2:A1625	+	47.00	116.63	351	protein kinase (EC 2.7.1.37), c
pir2:ORX082	+	47.00	115.64	397	rifin PRB0035c - malaria paras
pir2:T17588	+	47.00	115.47	375	protein kinase (EC 2.7.1.37), c
pir2:S46750	-	47.00	112.76	568	hyaluronoglucosaminidase-like p
pir2:A86190	+	47.00	109.47	856	aminopeptidase AAP1 (EC 3.4.11
	+	46.50	112.27	504	hypothetical protein [imported]

```
pir2:H86452 - 46.50 112.07 57.71 517 ! MUDRA-like transposase [impo
pir2:FB5649 - 46.50 111.74 57.78 539 ! hypothetical protein Z1543 {
pir2:T01865 - 46.50 111.15 57.91 580 ! hypothetical protein T24W8.2
pir2:FB4483 - 46.50 110.44 58.07 633 ! Mutator-like transposase [im
```

seq_name: pir2:JG0168

seq_documentation_block:

```
gob-5 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
C:Accession: JG0168
R:Komiyama, T.; Tanigawa, Y.; Hirohashi, S.
Biochem. Biophys. Res. Commun. 255, 347-351, 1999
A:Title: Cloning and identification of the gene gob-5, which is expressed in intestin
A:Reference number: JG0168; MUID:99160866
A:Accession: JG0168
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-913 <ROM>
A:Cross-references: DDBJ:AB016592
```

alignment_scores:

```
Quality: 90.00 Length: 19
Ratio: 4.737 Caps: 0
Percent Similarity: 100.000 Percent Identity: 68.421
```

alignment_block:

US-09-049-696-17 x JG0168

Align seg 1/1 to: JG0168 from: 1 to: 913

```
1 GCATTTCACATTTTAAATAATGTGGAAGTGGATAGAGAACTGCAGCT 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
892 GYIleHisValLeuLysileMetRlyTrpLeuGlyGluMetGlnVa 908
```

51 GTCATA 57

908 lThrLeu 910

seq_name: pir2:S40847

seq_documentation_block:

```
L-rhamnose isomerase (EC 5.3.1.14) - Escherichia coli
C:Species: Escherichia coli
C:Date: 19-May-1994 #sequence_revision 01-Sep-1995 #text_change 08-Oct-1999
C:Accession: S40847; B65196; B48649; S36640
R:Plunkett III, G.; Burland, V.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 21, 3391-3398, 1993
A:Title: Analysis of the Escherichia coli genome. III. DNA sequence of the region fro
A:Reference number: S40802; MUID:93347969
A:Accession: S40847
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-419 <PLU>
A:Cross-references: EMBL:L19201; NID:g304961; PIDN:AAB03036.1; PID:g305007
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1993
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: B65196
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-419 <BLAT>
A:Cross-references: GB:AE000465; GB:U00096; NID:g2367326; PIDN:AAC76885.1; PID:gl17903
A:Experimental source: strain K-12, substrain MG1655
R:Morales, P.; Egan, S.M.; Hidalgo, E.; Aguilar, J.
J. Bacteriol. 175, 5585-5594, 1993
A:Title: Sequencing and characterization of a gene cluster encoding the enzymes for L
A:Reference number: A48649; MUID:93374854
```

```

A:Accession: B48649
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-358, 'EL', 361-419 <MOR>
A:Cross-references: GB:X60472
A:Note: the sequence after residue 370 has been corrected for a frameshift error according to the EMBL Data Library, July 1991
R:Agullar, J.
A:Reference number: S36639
A:Accession: S36640
A:Molecule type: DNA
A:Residues: 1-358, 'EL', 361-365, 'P', 367-419 <AGU>
A:Cross-references: EMBL:X60472; NID:g396678; PIDN:CAA43002.1; PID:g396680
C:Genetics:
A:Gene: rhaA
C:Keywords: intramolecular oxidoreductase; isomerase

alignment_scores:
  Quality: 52.00 Length: 22
  Ratio: 3.250 Gaps: 1
  Percent Similarity: 72.727 Percent Identity: 54.545

alignment_block:
  US-09-049-696-17 x S40847 ..

Align seg 1/1 to: S40847 from: 1 to: 419

9 CATTTTAAAAATATCTGGAAGTGGATAGGAGAACTGCAGCTGTCATAG 58
|||||
114 HisPhelysAsnTrpValGlu.....TrpAlaLysAlaAsnG1 126

59 CCTAGGCTCAATTTT 74
:|||||
126 nLeuGlyLeuAspPhe 131

seq_name: pir2:E86079

seq_documentation_block:
L-rhamnose isomerase [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E86079
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.; Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca, Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E86079
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-419 <STO>
A:Cross-references: GB:AE005174; NID:g12518799; PIDN:AAG59097.1; GSPDB:GN00145; UWGP:Z54
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: rhaA

alignment_scores:
  Quality: 52.00 Length: 22
  Ratio: 3.250 Gaps: 1
  Percent Similarity: 72.727 Percent Identity: 54.545

alignment_block:
  US-09-049-696-17 x E86079 ..

Align seg 1/1 to: E86079 from: 1 to: 419

9 CATTTTAAAAATATCTGGAAGTGGATAGGAGAACTGCAGCTGTCATAG 58
|||||
114 HisPhelysAsnTrpValGlu.....TrpAlaLysAlaAsnG1 126

59 CCTAGGCTCAATTTT 74
:|||||
126 nLeuGlyLeuAspPhe 131

seq_name: pir2:E86079

seq_documentation_block:
L-rhamnose isomerase [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E86079
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.; Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca, Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E86079
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-419 <STO>
A:Cross-references: GB:AE005174; NID:g12518799; PIDN:AAG59097.1; GSPDB:GN00145; UWGP:Z54
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: rhaA

alignment_scores:
  Quality: 52.00 Length: 22
  Ratio: 3.250 Gaps: 1
  Percent Similarity: 72.727 Percent Identity: 54.545

alignment_block:
  US-09-049-696-17 x E86079 ..

Align seg 1/1 to: E86079 from: 1 to: 419

9 CATTTTAAAAATATCTGGAAGTGGATAGGAGAACTGCAGCTGTCATAG 58
|||||
114 HisPhelysAsnTrpValGlu.....TrpAlaLysAlaAsnG1 126

59 CCTAGGCTCAATTTT 74
:|||||
126 nLeuGlyLeuAspPhe 131

seq_name: pir2:E86079

seq_documentation_block:
L-rhamnose isomerase [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E86079
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.; Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca, Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E86079
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-419 <STO>
A:Cross-references: GB:AE005174; NID:g12518799; PIDN:AAG59097.1; GSPDB:GN00145; UWGP:Z54
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: rhaA

alignment_scores:
  Quality: 52.00 Length: 22
  Ratio: 3.250 Gaps: 1
  Percent Similarity: 72.727 Percent Identity: 54.545

alignment_block:
  US-09-049-696-17 x E86079 ..

Align seg 1/1 to: E86079 from: 1 to: 419

9 CATTTTAAAAATATCTGGAAGTGGATAGGAGAACTGCAGCTGTCATAG 58
|||||
114 HisPhelysAsnTrpValGlu.....TrpAlaLysAlaAsnG1 126

59 CCTAGGCTCAATTTT 74
:|||||
126 nLeuGlyLeuAspPhe 131

seq_name: pir2:E86079

seq_documentation_block:
L-rhamnose isomerase [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E86079
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.; Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca, Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E86079
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-419 <STO>
A:Cross-references: GB:AE005174; NID:g12518799; PIDN:AAG59097.1; GSPDB:GN00145; UWGP:Z54
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: rhaA

alignment_scores:
  Quality: 52.00 Length: 22
  Ratio: 3.250 Gaps: 1
  Percent Similarity: 72.727 Percent Identity: 54.545

alignment_block:
  US-09-049-696-17 x E86079 ..

Align seg 1/1 to: E86079 from: 1 to: 419

9 CATTTTAAAAATATCTGGAAGTGGATAGGAGAACTGCAGCTGTCATAG 58
|||||
114 HisPhelysAsnTrpValGlu.....TrpAlaLysAlaAsnG1 126

59 CCTAGGCTCAATTTT 74
:|||||
126 nLeuGlyLeuAspPhe 131

seq_name: pir2:E86079

seq_documentation_block:
L-rhamnose isomerase [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E86079
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.; Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca, Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E86079
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-419 <STO>
A:Cross-references: GB:AE005174; NID:g12518799; PIDN:AAG59097.1; GSPDB:GN00145; UWGP:Z54
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: rhaA

alignment_scores:
  Quality: 52.00 Length: 22
  Ratio: 3.250 Gaps: 1
  Percent Similarity: 72.727 Percent Identity: 54.545

alignment_block:
  US-09-049-696-17 x E86079 ..

Align seg 1/1 to: E86079 from: 1 to: 419

9 CATTTTAAAAATATCTGGAAGTGGATAGGAGAACTGCAGCTGTCATAG 58
|||||
114 HisPhelysAsnTrpValGlu.....TrpAlaLysAlaAsnG1 126

59 CCTAGGCTCAATTTT 74
:|||||
126 nLeuGlyLeuAspPhe 131

seq_name: pir2:E86079

seq_documentation_block:
L-rhamnose isomerase [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E86079
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.; Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca, Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E86079
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-419 <STO>
A:Cross-references: GB:AE005174; NID:g12518799; PIDN:AAG59097.1; GSPDB:GN00145; UWGP:Z54
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: rhaA

alignment_scores:
  Quality: 52.00 Length: 22
  Ratio: 3.250 Gaps: 1
  Percent Similarity: 72.727 Percent Identity: 54.545

alignment_block:
  US-09-049-696-17 x E86079 ..

Align seg 1/1 to: E86079 from: 1 to: 419

9 CATTTTAAAAATATCTGGAAGTGGATAGGAGAACTGCAGCTGTCATAG 58

```

```
US-09-049-696-17 x B84741
Align seg 1/1 to: B84741 from: 1 to: 740

3 CATTACATTTTAAATATGTTGGAAGTGGATAGGAGAACTGCAGCTGT 52
||||| :||| ||| ||| :||| :||| :|||
61 HisSerAspHisSerAsnGlyValTTPCysAspAsnSerThrGlyValVa 77
||||| :||| ||| ||| :||| :||| :|||
53 CAAATGCCCTAGGCGCTGAATTTTGTGCAGATAATAATAAATCAATCAT 102
||||| :||| ||| ||| :||| :||| :|||
77 lThrLysLeuGlnLeuAsnAlaCysLeuSerGlyThrLeuAsnProAsns 94
||||| :||| ||| ||| :||| :||| :|||
103 CC 104
94 er 94

seq_name: pir2:A64474

seq_documentation_block:
hypothetical protein MJ1394 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: A64474
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999
A:Accession: A64474
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-987 <BUIL>
A:Cross-references: GB:U67579; GB:L77117; NID:g1592037; PIDN:AAB99404.1; PID:g1592041; T
A:Map position: FOR1343894-1346857
A:Start codon: GTG

alignment_scores:
Quality: 52.00 Length: 13
Ratio: 4.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 53.846

alignment_block:
US-09-049-696-17 x A64474
Align seg 1/1 to: A64474 from: 1 to: 987

6 TCACATTTTAAATATGTTGGAAGTGGATAGGAGAACT 44
||||| :||| ||| ||| :||| :||| :|||
482 AsnHisTyrSerAsnTyrIleSerIleAspArgThr 494

seq_name: pir2:E84463

seq_documentation_block:
Mutator-like transposase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84463
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronan, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: E84463
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-616 <STO>
A:Cross-references: GB:A602093; NID:g4585938; PIDN:AAD25598.1; GSPDB:GN00139
C:Genetics:
```

```
A:Gene: At2g04950
A:Map position: 2

alignment_scores:
Quality: 51.50 Length: 27
Ratio: 2.575 Gaps: 1
Percent Similarity: 74.074 Percent Identity: 40.741

alignment_block:
US-09-049-696-17/rev x E84463
Align seg 1/1 to: E84463 from: 1 to: 616

99 AATGATTTTATTATTATCTGACAAAATTCAGCCCTAGGCTATTGACA 50
||||| :||| ||| ||| :||| :||| :|||
301 AsnAspLeuValPheValSerAspArgAlaAlaSerIleAlaSer...GI 316
||||| :||| ||| ||| :||| :||| :|||
49 GCTGCAGTTCTCTATCCACTTCCACATAAT 19
||||| :||| ||| ||| :||| :||| :|||
316 yLeuSerGlyAsnTyrProLeuAlaHisAsn 326
||||| :||| ||| ||| :||| :||| :|||

seq_name: pir1:VGBESM

seq_documentation_block:
glycoprotein B precursor - saimirine herpesvirus 1 (strain 11)
C:Species: saimirine herpesvirus 1
C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999
C:Accession: D34126
R:Albrecht, J.C.; Fleckenstein, B.
Virology 174, 533-542, 1990
A:Title: Structural organization of the conserved gene block of Herpesvirus saimiric
A:Reference number: A34126; MUID:90163221
A:Accession: D34126
A:Molecule type: DNA
A:Residues: 1-808 <ALB>
A:Cross-references: GB:M31122; NID:g331052; PIDN:AAA46164.1; PID:g331056
C:Superfamily: herpesvirus glycoprotein B
C:Keywords: glycoprotein; transmembrane protein
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-808/Product: glycoprotein B #status predicted <GLB>
F:675-692/Domain: transmembrane #status predicted <TM1>
F:702-722/Domain: transmembrane #status predicted <TM2>
F:30,158,239,251,285,331,344,355,361,471,532,569,587,598,727/Binding site: carbohydra

alignment_scores:
Quality: 50.50 Length: 36
Ratio: 2.196 Gaps: 2
Percent Similarity: 63.889 Percent Identity: 41.667

alignment_block:
US-09-049-696-17 x VGBESM
Align seg 1/1 to: VGBESM from: 1 to: 808

9 CATTTAAATTTATGGAAGTGGATAGGAGAACTGCAGCTGCAATAG 58
||||| :||| ||| ||| :||| :||| :|||
573 HisPheLysAsnTyrValHisValGlu.....ThrLeuProValAsnAs 587
||||| :||| ||| ||| :||| :||| :|||
59 C.....CTAGGCGCTGAATTTTGTGCAGATAAATA 87
||||| :||| ||| ||| :||| :||| :|||
587 nileSerThrLeuAspThrPheLeuAlaLeuAsnLeuThrPheIleGluA 604
||||| :||| ||| ||| :||| :||| :|||
88 AAATAAAT 95
||||| :||| ||| ||| :||| :||| :|||
604 snlleAsp 606
||||| :||| ||| ||| :||| :||| :|||

seq_name: pir2:D36828

seq_documentation_block:
orf13 protein - Autographa californica nuclear polyhedrosis virus
C:Species: Autographa californica nuclear polyhedrosis virus, AcMNPV
```

```
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: D36828; C72853
R:Brannagel, S.C.; Daniel, K.D.; Reilly, L.M.; Guarino, L.A.; Hong, T.; Summers, M.D.
Virology 191, 1003-1008, 1992
A:Title: Sequence, genomic organization of the EcoRI-A fragment of Autographa californica
VP8 of rotavirus.
A:Reference number: A44221; MUID:93079853
A:Accession: D36828
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-286 <BRA>
R:Avres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.
Virology 202, 586-605, 1994
A:Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.
A:Reference number: A72850; MUID:94303173
A:Accession: C72853
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-199, 'L', 201-286 <AYR>
A:Cross-references: GB:L22858; NID:g510708; PIDN:AAA66657.1; PID:g559096
C:Genetics:
C:Superfamily: viral apoptosis inhibitor IAP; RING finger homology

alignment_scores:
Quality: 50.00 Length: 32
Ratio: 2.273 Gaps: 2
Percent Similarity: 68.750 Percent Identity: 40.625

alignment_block:
US-09-049-696-17 x D36828 ..
Align seg 1/1 to: D36828 from: 1 to: 286
3 CATTCACATTTTAAATAT...GTGGAAGTGGATAGAGAACTGCACG 49
||||| |||:||||| :|||
33 HisSerPheGluAsnTyrProIleGlu.....AsnThrAlaPh 46
50 TGTCAATAGCGTGGCTCAATTTTGTGCAGATAAATAAATAAT 95
:|||||:||||| :||| :|||:|||||:
46 eIleAsnSerLeuIleValAsnGlyPheLysTyrAsnGlnValAsp 61

seq_name: pir2:T30149
seq_documentation_block:
Hypothetical protein C39H7.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T30149
R:Le, T.T.
submitted to the EMBL Data Library, May 1996
A:Description: The sequence of C. elegans cosmid C39H7.
A:Reference number: Z20744
A:Accession: T30149
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-391 <LET>
A:Cross-references: EMBL:U58754; PIDN:AA00671.1; GSPDB:GN00022; CESP:C39H7.4
A:Experimental source: strain Bristol N2; clone C39H7
C:Genetics:
A:Gene: CESP:C39H7.4
A:Map position: 4
A:Introns: 39/1; 275/3; 308/2; 365/1

alignment_scores:
Quality: 50.00 Length: 32
Ratio: 2.273 Gaps: 0
Percent Similarity: 68.750 Percent Identity: 28.125

alignment_block:
US-09-049-696-17 x T30149 ..
Align seg 1/1 to: T30149 from: 1 to: 391
9' CATTTTAAATATATGCAAGTGGATAGAGAACTGCAGCTGTCAATAG 58
|||||:||||| ||| :|||:|||||:
280 HisPheSerAsnThrValAlaIleLeuIleSerIleThrValIleValPh 296
59 CCTAGGGCTGAATTTTGTGCAGATAAATAAATAATCATTCATCC 104
:|||||:||||| :|||:|||||:
296 eAlaAlaIleSerPheGlnAspValAsnLysValAsnHisLysThr 311

seq_name: pir2:T03377
seq_documentation_block:
homeotic protein HOX1B - maize
C:Species: Zea mays (maize)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T03377
R:Deberlaecker, B.; Klinge, B.; Werr, W.
Plant Cell 8, 349-362, 1996
A:Title: Ectopic expression of the maize homeobox genes ZmHox1a or ZmHox1b causes ple
A:Reference number: Z14909; MUID:96361874
A:Accession: T03377
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-692 <OEB>
A:Cross-references: EMBL:X92428; NID:g1648930; PIDN:CAA63156.1; PID:g1648931
C:Genetics:
A:Gene: Hox1b
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

alignment_scores:
Quality: 50.00 Length: 31
Ratio: 2.174 Gaps: 1
Percent Similarity: 74.194 Percent Identity: 35.484

alignment_block:
US-09-049-696-17 x T03377 ..
Align seg 1/1 to: T03377 from: 1 to: 692
3 CATTCACATTTTAAAT...TATGTGAAGTGGATAGAGAACTGC 46
|||||:|||||:||||| :||| :|||:|||||:
577 HisGluHisPheLysThrGlnProTyrProSerArgSerLeuLysGluSe 593
47 AGCTGTCAATAGCCCTAGCGCTGAATTTTGTGCAGATAAATAA 89
:|||||:|||||:||||| :|||:|||||:
593 rLeuAlaGluGluLeuGlyLeuThrPheHisGlnValAsnArg 607

seq_name: pir2:B81711
seq_documentation_block:
conserved hypothetical protein TC0355 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: B81711
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: B81711
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-558 <TET>
A:Cross-references: GB:AE002303; GB:AE002160; NID:g7190392; PIDN:AAF39216.1; PID:g719
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0355
```

```
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Nov-2000
C:Accession: D36828; C72853
R:Brannagel, S.C.; Daniel, K.D.; Reilly, L.M.; Guarino, L.A.; Hong, T.; Summers, M.D.
Virology 191, 1003-1008, 1992
A:Title: Sequence, genomic organization of the EcoRI-A fragment of Autographa californica
VP8 of rotavirus.
A:Reference number: A44221; MUID:93079853
A:Accession: D36828
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-286 <BRA>
R:Avres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.
Virology 202, 586-605, 1994
A:Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.
A:Reference number: A72850; MUID:94303173
A:Accession: C72853
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-199, 'L', 201-286 <AYR>
A:Cross-references: GB:L22858; NID:g510708; PIDN:AAA66657.1; PID:g559096
C:Genetics:
C:Superfamily: viral apoptosis inhibitor IAP; RING finger homology

alignment_scores:
Quality: 50.00 Length: 32
Ratio: 2.273 Gaps: 2
Percent Similarity: 68.750 Percent Identity: 40.625

alignment_block:
US-09-049-696-17 x D36828 ..
Align seg 1/1 to: D36828 from: 1 to: 286
3 CATTCACATTTTAAATAT...GTGGAAGTGGATAGAGAACTGCACG 49
||||| |||:||||| :|||
33 HisSerPheGluAsnTyrProIleGlu.....AsnThrAlaPh 46
50 TGTCAATAGCGTGGCTCAATTTTGTGCAGATAAATAAATAAT 95
:|||||:||||| :||| :|||:|||||:
46 eIleAsnSerLeuIleValAsnGlyPheLysTyrAsnGlnValAsp 61

seq_name: pir2:T30149
seq_documentation_block:
Hypothetical protein C39H7.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T30149
R:Le, T.T.
submitted to the EMBL Data Library, May 1996
A:Description: The sequence of C. elegans cosmid C39H7.
A:Reference number: Z20744
A:Accession: T30149
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-391 <LET>
A:Cross-references: EMBL:U58754; PIDN:AA00671.1; GSPDB:GN00022; CESP:C39H7.4
A:Experimental source: strain Bristol N2; clone C39H7
C:Genetics:
A:Gene: CESP:C39H7.4
A:Map position: 4
A:Introns: 39/1; 275/3; 308/2; 365/1

alignment_scores:
Quality: 50.00 Length: 32
Ratio: 2.273 Gaps: 0
Percent Similarity: 68.750 Percent Identity: 28.125

alignment_block:
US-09-049-696-17 x T30149 ..
Align seg 1/1 to: T30149 from: 1 to: 391
9' CATTTTAAATATATGCAAGTGGATAGAGAACTGCAGCTGTCAATAG 58
|||||:||||| ||| :|||:|||||:
280 HisPheSerAsnThrValAlaIleLeuIleSerIleThrValIleValPh 296
59 CCTAGGGCTGAATTTTGTGCAGATAAATAAATAATCATTCATCC 104
:|||||:||||| :|||:|||||:
296 eAlaAlaIleSerPheGlnAspValAsnLysValAsnHisLysThr 311

seq_name: pir2:T03377
seq_documentation_block:
homeotic protein HOX1B - maize
C:Species: Zea mays (maize)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T03377
R:Deberlaecker, B.; Klinge, B.; Werr, W.
Plant Cell 8, 349-362, 1996
A:Title: Ectopic expression of the maize homeobox genes ZmHox1a or ZmHox1b causes ple
A:Reference number: Z14909; MUID:96361874
A:Accession: T03377
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-692 <OEB>
A:Cross-references: EMBL:X92428; NID:g1648930; PIDN:CAA63156.1; PID:g1648931
C:Genetics:
A:Gene: Hox1b
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

alignment_scores:
Quality: 50.00 Length: 31
Ratio: 2.174 Gaps: 1
Percent Similarity: 74.194 Percent Identity: 35.484

alignment_block:
US-09-049-696-17 x T03377 ..
Align seg 1/1 to: T03377 from: 1 to: 692
3 CATTCACATTTTAAAT...TATGTGAAGTGGATAGAGAACTGC 46
|||||:|||||:||||| :||| :|||:|||||:
577 HisGluHisPheLysThrGlnProTyrProSerArgSerLeuLysGluSe 593
47 AGCTGTCAATAGCCCTAGCGCTGAATTTTGTGCAGATAAATAA 89
:|||||:|||||:||||| :|||:|||||:
593 rLeuAlaGluGluLeuGlyLeuThrPheHisGlnValAsnArg 607

seq_name: pir2:B81711
seq_documentation_block:
conserved hypothetical protein TC0355 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: B81711
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: B81711
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-558 <TET>
A:Cross-references: GB:AE002303; GB:AE002160; NID:g7190392; PIDN:AAF39216.1; PID:g719
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0355
```


A; Cross-references: GB:X62052; NID:G396598; PIDN:CAA43980.1; PID:G396599
P; Blumberg, B.; Wright, C.V.E.; De Robertis, E.M.; Cho, K.W.Y.
Science, 253, 194-196, 1991
A; Title: Organizer-specific homeobox genes in *Xenopus laevis* embryos.
A; Reference number: A40856; MUID:91305940
A; Accession: C40856
A; Status: preliminary; not compared with conceptual translation
A; Molecule type: mRNA
A; Residues: 126-141, 'O', 143-185 <BLU>
A; Cross-references: GB:M63873
C; Superfamily: unassigned homeobox proteins; homeobox homology
C; Keywords: DNA binding; homeobox; nucleus; transcription regulation
F; 127-183/Domain: Homeobox homology <HOX>

```

alignment_block:
  US-09-049-696-17 x A40357  ..
  Align_seg 1/1  to: A40357  from: 1  to: 272
      9  CATTTAAAAAATTATTCGGAAGTGGATGAGAGAACTGCAGCTGTCAATAG  58
         |||||.....:|||||  |||||
      145 HisPheAsnLysTyrlLeuthrArgAlaArglleGluIleAlaAsnSe  162
         .....:|||||  |||||
      159  CCTAGGCGTGAATTTTGTGCAGATA  83
         ||||  |||||  |||||
      162 rLeuGlnLeuAsnAspThrGlnVal  170
         .....:|||||  |||||
seq name: pir2.T41772

```

```

seq_documentation_block:
  IAP1 orf27 - Bombyx mori nuclear polyhedrosis virus (isolate T3)
  C:Species: Bombyx mori nuclear polyhedrosis virus, BmSNPV
  A:Variety: isolate T3
  C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Sep-2000
  C:Accession: T41772
  R:Gomi, S.; Majima, K.; Maeda, S.
  J. Gen. Virol. 80, 1323-1337, 1999
  A:title: Sequence analysis of the genome of Bombyx mori nucleopolyhedrovirus
  A:Reference number: Z22020; MUID:99281911
  A:Accession: T41772
  A:Status: preliminary; translated from GB/EMBL/DBJ
  A:Molecule type: DNA
  A:Residues: 1-292 <RAM>
  A:Cross-references: EMBL:L33180; NID:g3745835; PIDN:AAC63701.1; PID:g3745854
  A:Experimental source: isolate T3

```

```

C:Superfamily: viral apoptosis inhibitor IAP; RING finger homology
alignment_scores:
    Quality: 49.00      Length: 31
    Ratio: 2.333       Gaps: 1
    Percent Similarity: 67.742    Percent Identity: 38.710
alignment_block:
US-09-949-696-17 x T41772 ..
Align seg 1/1 to: T41772 from: 1 to: 292
33 CAATCACATTTTAAAAATTATGTGGAACTGGATAGGAGAACTGCAGCTGT 52
||||| : ||||| ||||| : ||||| :
33 HisGlySerPhegluAsnTyrProIleValasn.....ThrAlaPheII 47
33 CAATAGCCTAGGCTGAATTTTGTCTCAGATAAATAAATAAAT 95
||||| : ||||| : ||||| : ||||| :
47 easnSerLeuIIleValIasnGlyPheLysTyrAsnGlnValasp 61

```

This Page Blank (uspto)

OM of: US-09-049-696-17 to: Issued_Patents_AA:* out_format : pfs

Date: Mar 30, 2002 2:33 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame2_n2p.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09049696/runat_28032002_145238_2012/app_query.fasta_1.12579
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=ra1 -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOAPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DLOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=EXT -MINLEN=0 -MAXLEN=200000000
-USER=US09049696@cgn1_1109 -NCPU=6 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-17
Query length: 106
Database: Issued_Patents_AA:*
Database sequences: 212252
Database length: 22503292
Search time (sec): 362.110000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-804-439A-14		50.50	120.47	3.64	808
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-720-229-14		50.50	120.47	3.64	808
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-08-468-576B-12		49.00	118.73	6.08	604
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-08-468-576B-12		49.00	118.73	6.08	604
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-468-577B-12		49.00	118.73	6.08	604
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-08-742-440A-6		47.00	116.45	12.07	408
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-08-878-474-3		46.00	115.80	16.82	318
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-08-742-440A-3		46.00	113.50	17.65	407
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-870-529-2		46.00	105.32	20.96	979
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-633-933A-15		45.00	114.16	23.94	276
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-844-188-15		45.00	114.16	23.94	276
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-633-933A-13		45.00	114.09	23.97	278
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-844-188-13		45.00	114.09	23.97	278
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-466-343D-9		45.00	112.10	24.99	344
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-461-244-3		45.00	112.02	25.04	347
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-450-393A-4		45.00	111.68	25.22	360
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-08-446-669-4		45.00	111.68	25.22	360
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-09-045-383-50		45.00	111.68	25.22	360
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-450-393A-2		45.00	111.32	25.41	374
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-08-446-669-2		45.00	111.32	25.41	374
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-844-188-38		45.00	111.10	25.52	393
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-844-188-43		45.00	111.10	25.52	383
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-08-826-267-2		45.00	102.36	37.84	2860
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-08-392-625-20		44.50	100.76	37.19	990
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-08-466-961A-20		44.50	100.76	37.19	990
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-08-645-193B-15		44.50	100.76	37.19	990
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-804-439A-93		44.00	117.85	30.45	135
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-720-229-93		44.00	117.85	30.45	135
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-09-067-089-2		44.00	108.87	36.78	354
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-09-045-583-51		44.00	108.71	36.90	360
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-594-031-90		44.00	104.59	40.24	560
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-08-456-647B-20		43.50	99.17	52.87	854
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-08-237-401A-20		43.50	99.17	52.87	854
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-08-549-004A-6		43.00	123.60	37.11	53
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-09-051-982A-6		43.00	123.60	37.11	53
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-370-583-2		43.00	105.84	53.89	356
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-783-889A-2		43.00	105.84	53.89	356
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-09-091-058-2		43.00	103.06	57.13	480

/cgn2_6/ptodata/2/iaa/5A_COMB.pcp:US-08-220-151-6 + 43.00 97.06 64.80 913
/cgn2_6/ptodata/2/iaa/5A_COMB.pcp:US-08-413-118-6 + 43.00 97.06 64.80 913
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-473-446-6 + 43.00 97.06 64.80 913
/cgn2_6/ptodata/2/iaa/6B_COMB.pcp:US-09-232-468A-2 + 43.00 97.06 64.80 913
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-09-232-468A-2 + 43.00 97.06 64.80 913
/cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-804-439A-14

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pcp:US-08-804-439A-14

seq_documentation_block:

Sequence 14, Application US/08804439A
Patent No. 6015565
GENERAL INFORMATION:
APPLICANT: Rose, Timothy M.
APPLICANT: Bosch, Marnix L.
TITLE OF INVENTION: GLYCOPROTEIN B OF THE RFHV/KSHV
SUBFAMILY OF HERPES VIRUSES
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Ste 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804.439A
FILING DATE: February 21, 1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09176/004001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 678-5070
TELEFAX: (619) 678-5099
TELEX:

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:
LENGTH: 808 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-804-439A-14

alignment_scores:

Quality: 50.50 Length: 36
Ratio: 2.196 Gaps: 2
Percent Similarity: 63.889 Percent Identity: 41.667

alignment_block:

US-09-049-696-17 x US-08-804-439A-14

Align seg 1/1 to: US-08-804-439A-14 from: 1 to: 808

9 CATTTTAAATATATGGAGTGGATAGGAGAACTGCAGCTGCTAATAG 58
|||||
573 HisPheLysAsnTyrValHisValGlu.....ThrLeuProValAsnAs 587
59 C.....CTAGGCGTGAATTTTGTGCAGATAAATA 87
587 nileSerThrLeuAspThrPheLeuAlaLeuAsnLeuThrPheileGluA 604
88 AATAAAT 95
604 snileasp 606

```
seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-720-229-14
seq_documentation_block:
; Sequence 14, Application US/08720229
; Patent No. 6022542
; GENERAL INFORMATION:
; APPLICANT: Rose, Timothy M.
; APPLICANT: Bosch, Marix L.
; APPLICANT: Strand, Kurt
; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RHV/KSHV
; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES
; NUMBER OF SEQUENCES: 100
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/720,229
; FILING DATE: 26-SEP-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Schiff, J. Michael
; REGISTRATION NUMBER: 40,253
; REFERENCE/DOCKET NUMBER: 29938-20002.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 808 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-720-229-14

alignment_scores:
Quality: 50.50 Length: 36
Ratio: 2.196 Gaps: 2
Percent Similarity: 63.889 Percent Identity: 41.667

alignment_block:
US-09-049-696-17 x US-08-720-229-14 ..
Align seg 1/1 to: US-08-720-229-14 from: 1 to: 808

9 CATTTTAAATATATGGAAGTGGATAGAGAACTGCAGCTGCAATAG 58
|||||
573 HisPheLysAsnTyrValHisValGlu.....ThrLeuProValAsnAs 587
59 C.....CTAGGGCTGAATTTTGTGCAGATAAATA 87
:|||||
587 nileSerThrLeuAspThrPheLeuAlaLeuAsnLeuThrPheIleGluA 604
88 AATAAAT 95
:|||||
604 snleasp 606

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-468-576B-12
seq_documentation_block:
; Sequence 12, Application US/08468576B
```

```
; Patent No. 5955345
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,576B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.7-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-468-576B-12

alignment_scores:
Quality: 49.00 Length: 20
Ratio: 3.062 Gaps: 0
Percent Similarity: 80.000 Percent Identity: 40.000

alignment_block:
US-09-049-696-17 x US-08-468-576B-12 ..
Align seg 1/1 to: US-08-468-576B-12 from: 1 to: 604

30 GTGGATAGGAGAACTGCAGCTGTCGAATAGCCTAGGCTGAATTTTGTC 79
|||||
523 ValAspGlyLysLysSerSerIleAsnAsnMetGlyLeuAsnSerCysAr 538
80 GATAAATAAA 89
:|||||
538 gAsnGluLys 541

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-468-579B-12
seq_documentation_block:
```

```
; Sequence 12, Application US/08468579B
; Patent No. 5981700
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,579B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.5-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-468-579B-12
```

```
alignment_scores:
  Quality: 49.00      Length: 20
  Ratio: 3.062       Gaps: 0
  Percent Similarity: 80.000  Percent Identity: 40.000

alignment_block:
  US-09-049-696-17 x US-08-468-579B-12  ..

  Align seg 1/1 to: US-08-468-579B-12 from: 1 to: 604

      30 GTGGATAGGAGAACTGCAGCTGTCATAGCCTAGGCTGAATTTTGTCA 79
          ||||| : : : : : : : : : : : : : : : : : : : : : :
      522 ValAspGlyLysLysSerSerIleAsnAsnMetGlyLeuAsnSerCysAr 538

      80 GATAAATAAA 89
          : : : : :
      538 gAsnGluLys 541
```

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pap:US-08-468-577B-12

```
seq_documentation_block:
; Sequence 12, Application US/08468577B
; Patent No. 6001804
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,577B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,276
; FILING DATE: 05-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.8-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-468-577B-12
```

```
alignment_scores:
  Quality: 49.00      Length: 20
  Ratio: 3.062       Gaps: 0
  Percent Similarity: 80.000  Percent Identity: 40.000

alignment_block:
  US-09-049-696-17 x US-08-468-577B-12  ..

  Align seg 1/1 to: US-08-468-577B-12 from: 1 to: 604

      30 GTGGATAGGAGAACTGCAGCTGTCATAGCCTAGGCTGAATTTTGTCA 79
          ||||| : : : : : : : : : : : : : : : : : : : : : :
      522 ValAspGlyLysLysSerSerIleAsnAsnMetGlyLeuAsnSerCysAr 538

      80 GATAAATAAA 89
          : : : : :
      538 gAsnGluLys 541
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pap:US-08-742-440A-6

```
seq_documentation_block:
; Sequence 6, Application US/08742440A
; Patent No. 5892014
; GENERAL INFORMATION:
; APPLICANT: Coughlin, Shaun
; APPLICANT: Ishihari, Hiroaki
; APPLICANT: Connolly, Andrew
; TITLE OF INVENTION: Protease Activated Receptor
; NUMBER OF SEQUENCES: 23
; TITLE OF INVENTION: 3 and Uses Thereof
; CORRESPONDENCE ADDRESS:
; STREET: 285 Hamilton Avenue, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742,440A
; FILING DATE: 30-OCT-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: UCAL/060PAT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3400
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
;
US-08-742-440A-6

alignment_scores:
Quality: 47.00 Length: 28
Ratio: 2.474 Gaps: 0
Percent Similarity: 67.857 Percent Identity: 32.143

alignment_block:
US-09-049-696-17/rev x US-08-742-440A-6
..
Align seg 1/1 to: US-08-742-440A-6 from: 1 to: 408

95 ATTTATTTTATTTATCTGACAAAATTCAGCCCTAGGCTATTGACAGCTG 46
:::|||||:::
355 LeuTyrrPheIleTyrrLeulleAlaLeuCySerLeuGlySerLeuAsnSerCy 371
| :::::
45 CAGTTCCTCATCCACTCCACATAATTTTAA 12
371 sLeuAspPropheLeuTyrrPheLeuMetSerLys 382

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-878-474-3

seq_documentation_block:
; Sequence 3, Application US/08878474
; Patent No. 6135232
; GENERAL INFORMATION:
; APPLICANT: De Robertis, Edward M.
```

```
APPLICANT: Bouwmeester, Tewis
TITLE OF INVENTION: Endoderm, Cardiac and Neural Inducing
TITLE OF INVENTION: Factors
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Majestic, Parsons, Siebert & Hsue
STREET: Four Embarcadero Center, Suite 1100
CITY: San Francisco
STATE: California
COUNTRY: U.S.A.
ZIP: 94111-4106
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,474
FILING DATE: 18-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/020,150
FILING DATE: 20-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Siebert, J. Suzanne
REGISTRATION NUMBER: 28,758
REFERENCE/DOCKET NUMBER: 3100.002US1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/248-5500
TELEFAX: 415/362-5418
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-878-474-3

alignment_scores:
Quality: 46.00 Length: 11
Ratio: 4.600 Gaps: 0
Percent Similarity: 90.909 Percent Identity: 63.636

alignment_block:
US-09-049-696-17 x US-08-878-474-3
..
Align seg 1/1 to: US-08-878-474-3 from: 1 to: 318

23 TGTGGAAGTGGATAGGAGAACTGCAGCTGTCAA 55
|||||:::
171 CysGlySerGlyArgGluHisCysLysCysLys 181

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-742-440A-3

seq_documentation_block:
; Sequence 3, Application US/08742440A
; Patent No. 5892014
; GENERAL INFORMATION:
; APPLICANT: Coughlin, Shaun
; APPLICANT: Ishihari, Hiroaki
; APPLICANT: Connolly, Andrew
; TITLE OF INVENTION: Protease Activated Receptor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Avenue, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

REFERENCE/DOCKET NUMBER: MA703
TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 276 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: protein
; US-08-633-993A-15
;
alignment_scores:
    Quality: 45.00      Length: 29
    Ratio: 2.250        Gaps: 2
    Percent Similarity: 68.966      Percent Identity: 37.931
;
alignment_block:
    US-09-049-696-17 x US-08-633-993A-15 ..
;
Align seg 1/1 to: US-08-633-993A-15 from: 1 to: 276
;
15 AAAAATTAT.....GTGGAAGTGGATAGGAGAACTGCAGC 49
   |||:||||
210 LysSerTyrAlaTyrGluTrpGlyThrGluLeuAspGlnLysThrThrIl 226
;
50 TGTCAATAGCCTAGGCTGGAATTTTGTGCAGATAAAT 86
   |||:||||
226 eileAsnThrLeuGlyPhe.....GlnIleAsn 235
;
seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-633-993A-15
;
seq_documentation_block:
; Sequence 15, Application US/08844188
; Patent No. 6127180
; GENERAL INFORMATION:
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Knuth, Mark
; APPLICANT: Pollard, Michael R.
; APPLICANT: Cardineau, Guy
; APPLICANT: Schwab, George E.
; TITLE OF INVENTION: Pesticidal Toxins
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/844,188
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/633,993
; FILING DATE: 19-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanders, Jay M.
; REGISTRATION NUMBER: 39,355
; REFERENCE/DOCKET NUMBER: MA-703C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 276 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: protein
; US-08-844-188-15
;
alignment_scores:
    Quality: 45.00      Length: 29
    Ratio: 2.250        Gaps: 2
    Percent Similarity: 68.966      Percent Identity: 37.931
;
alignment_block:
    US-09-049-696-17 x US-08-844-188-15 ..
;
Align seg 1/1 to: US-08-844-188-15 from: 1 to: 276
;
15 AAAAATTAT.....GTGGAAGTGGATAGGAGAACTGCAGC 49
   |||:||||
210 LysSerTyrAlaTyrGluTrpGlyThrGluLeuAspGlnLysThrThrIl 226
;
50 TGTCAATAGCCTAGGCTGGAATTTTGTGCAGATAAAT 86
   |||:||||
226 eileAsnThrLeuGlyPhe.....GlnIleAsn 235
;
seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-633-993A-13
;
seq_documentation_block:
; Sequence 13, Application US/08633993A
; Patent No. 6083499
; GENERAL INFORMATION:
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Schnepf, H. Ernest
; APPLICANT: Knuth, Mark
; APPLICANT: Pollard, Michael R.
; APPLICANT: Cardineau, Guy
; APPLICANT: Schwab, George E.
; TITLE OF INVENTION: Pesticidal Toxins
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,993A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA703
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: protein
; US-08-633-993A-13

```



```
alignment_scores:
  Quality: 45.00      Length: 29
  Ratio: 2.250       Gaps: 2
  Percent Similarity: 68.966   Percent Identity: 37.931

alignment_block:
US-09-049-696-17 x US-08-633-993A-13
..
Align seg 1/1 to: US-08-633-993A-13 from: 1 to: 278

15 AAAAATTAT.....GTGGAAGTGGATAGGAGAACTGCAGC 49
||||:||||
212 LysSerTyrThrTyrGluTrpGlyThrGluLeuAspGlnLysThrThr 228
||||:||||
50 TGTCAATAGCCTAGGCTGAATTTTGTTCAGATAAAT 86
||||:||||
228 eileAsnThrLeuGlyPhe.....GlnIleAsn 237
||||:||||

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pap:US-08-844-188-13

seq_documentation_block:
; Sequence 13, Application US/08844188
; Patent No. 6127180
; GENERAL INFORMATION:
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Schepf, H. Ernest
; APPLICANT: Knuth, Mark
; APPLICANT: Pollard, Michael R.
; APPLICANT: Cardineau, Guy
; APPLICANT: Schwab, George E.
; TITLE OF INVENTION: Pesticidal Toxins
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32608-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/844,188
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/633,993
; FILING DATE: 19-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanders, Jay M.
; REGISTRATION NUMBER: 39,355
; REFERENCE/DOCKET NUMBER: MA-703C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-844-188-13

alignment_scores:
  Quality: 45.00      Length: 29
  Ratio: 2.250       Gaps: 2
  Percent Similarity: 68.966   Percent Identity: 37.931
```

```
alignment_block:
US-09-049-696-17 x US-08-844-188-13
..
Align seg 1/1 to: US-08-844-188-13 from: 1 to: 278

15 AAAAATTAT.....GTGGAAGTGGATAGGAGAACTGCAGC 49
||||:||||
212 LysSerTyrThrTyrGluTrpGlyThrGluLeuAspGlnLysThrThr 228
||||:||||
50 TGTCAATAGCCTAGGCTGAATTTTGTTCAGATAAAT 86
||||:||||
228 eileAsnThrLeuGlyPhe.....GlnIleAsn 237
||||:||||

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pap:US-08-466-343D-9

seq_documentation_block:
; Sequence 9, Application US/08466343D
; Patent No. 6025154
; GENERAL INFORMATION:
; APPLICANT: LI, YI
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING HUMAN G-PROTEIN
; TITLE OF INVENTION: CHEMOKINE RECEPTOR HDGNR10 (AS AMENDED)
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,343D
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.1150000/EKS/KLM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-466-343D-9

alignment_scores:
  Quality: 45.00      Length: 29
  Ratio: 2.143       Gaps: 0
  Percent Similarity: 72.414   Percent Identity: 31.034

alignment_block:
US-09-049-696-17 x US-08-466-343D-9
..
Align seg 1/1 to: US-08-466-343D-9 from: 1 to: 344

16 TCACATTTAAAAAATATGTGGAAGTGGATAGGAGAACTGCAGCTGTC 55
||||:||||
256 SerAsnCysGluSerThrSerGlnLeuAspGlnAlaThrGlnValThrGl 274
||||:||||
50 TAGCCTAGGCTCAATTTTGTTCAGATAAATAAATA 92
||||:||||
```

```
274 uThrLeuGlyMetThrHisCysCysIleAsnProIle 286
seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-461-244-3
seq_documentation_block:
; Sequence 3, Application US/08461244
; Patent No. 5776729
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel R.
; APPLICANT: Yi, Li
; APPLICANT: Ruben, Steven M.
; APPLICANT: Rosen, Craig A.
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HGBER32
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,244
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-445
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 347 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-461-244-3

alignment_scores:
Quality: 45.00 Length: 29
Ratio: 2.143 Gaps: 0
Percent Similarity: 72.414 Percent Identity: 31.034

alignment_block:
US-09-049-696-17 x US-08-461-244-3 ..
Align seg 1/1 to: US-08-461-244-3 from: 1 to: 347

6 TCACATTTTAAATAATTATGTGGAACTGGATAGGAGAACTGCAGCTGTCAA 55
||||| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
262 SerAsnCysGluSerThrSerGlnLeuAspGlnAlaThrGlnValThrGI 278
56 TAGCGTAGGGCTGAATTTTGTCTCAGATAATAATAAATA 92
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
278 uThrLeuGlyMetThrHisCysCysIleAsnProIle 290
```

/SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW53549 +	46.00	118.57	40.13
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAG72283 +	46.00	118.47	40.21
/SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW41251 +	46.00	118.35	40.32
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAG72719 +	46.00	118.15	40.51
/SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW51405 -	46.00	116.95	41.62

```
seq_name: /SIDS2/qcadata/qeneseq/qeneseqp/AA2001.DAT:AAAG73854
```

OS	
XX	homio septiens.
XX	
PN	WO200122920-A2.
XX	
PD	05-APR-2001.
XX	
PF	28-SEP-2000; 2000WO-US26524.
XX	
PR	29-SEP-1999; 99US-0157137.
PR	03-NOV-1999; 99US-0163280.
XX	
XX	{HUMA-) HUMAN GENOME SCI INC.
PA	
XX	Ruben SM, Barash SC, Birse CE, Rosen CA;
PI	
PT	WPT: 2001-235357/24.
DR	N-PSTDB; AAH33285.
DR	
XX	Nucleic acids encoding 4277 human colon cancer-associated polypeptides
PT	useful for preventing, diagnosing and/or treating colorectal cancers -
PT	
PS	Claim 11: Page 6416-6419; 9803pp. English.
XX	

PN WO200122920-A2.
YY

05-APR-2001
XX
XX

FD 03-APR-2001.
XX

28-SEP-2000: 2000WO-IIS26524

XX 14
20 SEP 2000, 2000W 0520JZ4.

29-SEP-1999: 99US-0157137.

score_list::	sequence	Strd Orig	zScore	EScore	Len	! Documentation	RA
	/SID52/gcgdata/geneseq/A22001.DAT: AAG73854				107.00	300.22	552
	/SID52/gcgdata/geneseq/A22001.DAT: AAG75614				107.00	295.93	869
	/SID52/gcgdata/geneseq/A22001.DAT: AAB74824				107.00	295.45	914
	/SID52/gcgdata/geneseq/A22001.DAT: AAM24514				107.00	295.45	914
	/SID52/gcgdata/geneseq/A22001.DAT: AAB73716				107.00	295.45	914
	/SID52/gcgdata/geneseq/A22001.DAT: AAB74733				107.00	295.45	914
	/SID52/gcgdata/geneseq/A22001.DAT: AAG75474				107.00	295.34	925
	/SID52/gcgdata/geneseq/A22001.DAT: AAB74822				90.00	243.32	913
	/SID52/gcgdata/geneseq/A22001.DAT: AAB73715				90.00	243.32	913
	/SID52/gcgdata/geneseq/A22001.DAT: AAG72075				53.00	139.44	331
	/SID52/gcgdata/geneseq/A22001.DAT: AAG72874				52.00	136.58	315
	/SID52/gcgdata/geneseq/A22001.DAT: AAG72677				52.00	136.85	324
	/SID52/gcgdata/geneseq/A22001.DAT: AAB92584				51.00	129.05	626
	/SID52/gcgdata/geneseq/A22001.DAT: AAB67757				51.00	124.33	855
	/SID52/gcgdata/geneseq/A22001.DAT: AAB95509				50.00	120.83	832
	/SID52/gcgdata/geneseq/A22001.DAT: AAG93419				50.00	128.17	405
	/SID52/gcgdata/geneseq/A22001.DAT: AAG71938				49.00	128.17	298
	/SID52/gcgdata/geneseq/A22001.DAT: AAG19356				49.00	127.29	327
	/SID52/gcgdata/geneseq/A22001.DAT: AAR06463				49.00	121.48	604
	/SID52/gcgdata/geneseq/A22001.DAT: AAY49848				49.00	121.48	604
	/SID52/gcgdata/geneseq/A22001.DAT: AAY33354				49.00	121.48	604
	/SID52/gcgdata/geneseq/A22001.DAT: AAW80480				49.00	121.48	604
	/SID52/gcgdata/geneseq/A22001.DAT: AAY49322				49.00	121.48	604
	/SID52/gcgdata/geneseq/A22001.DAT: AAG39081				49.00	121.31	615
	/SID52/gcgdata/geneseq/A22001.DAT: AAG03253				48.00	137.20	83
	/SID52/gcgdata/geneseq/A22001.DAT: AAG72258				47.50	124.79	262
	/SID52/gcgdata/geneseq/A22001.DAT: AAM14741				47.00	140.14	44
	/SID52/gcgdata/geneseq/A22001.DAT: AAM27164				47.00	140.14	44
	/SID52/gcgdata/geneseq/A22001.DAT: AAM02463				47.00	140.14	44
	/SID52/gcgdata/geneseq/A22001.DAT: AAW51406				47.00	119.89	374
	/SID52/gcgdata/geneseq/A22001.DAT: AAB18149				47.00	119.86	375
	/SID52/gcgdata/geneseq/A22001.DAT: AAY06207				47.00	115.95	567
	/SID52/gcgdata/geneseq/A22001.DAT: AAW63675				47.00	110.44	1015
	/SID52/gcgdata/geneseq/A22001.DAT: AAB73618				47.00	110.44	1015
	/SID52/gcgdata/geneseq/A22001.DAT: AAB41442				46.00	127.34	123
	/SID52/gcgdata/geneseq/A22001.DAT: AAR75902				46.00	122.02	216
	/SID52/gcgdata/geneseq/A22001.DAT: AAR75904				46.00	122.02	216
	/SID52/gcgdata/geneseq/A22001.DAT: AAM24365				46.00	120.59	383
	/SID52/gcgdata/geneseq/A22001.DAT: AAG71637				46.00	118.60	310
	/SID52/gcgdata/geneseq/A22001.DAT: AAG72594				46.00	118.60	310

(HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Barash SC, Birse CE, Rosen CA;

WPT: 2001-235357/24.

N+PSDB; AAG73854.

Nucleic acids encoding 4277 human colon cancer-associated polypeptides - useful for preventing, diagnosing and/or treating colorectal cancers -

Claim 11; Page 6416-6419; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with a patient's genome that affect the activity of P by expressing P in inactive proteins or to supplement the patient's own production of P. Additionally, N may be used to produce the colon cancer-associated P, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37299 represent sequences used in the exemplification of the present invention.

N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922.

Sequence 552 AA;

alignment_scores:

Quality: 107.00 Length: 20

Ratio: 5.350 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-17 x AAG73854

```
alignment_block:
. US-09-049-696-17 x AAG73854
```

Align seg 1/1 to: AAG73854 from: 1 to: 552

1 GGCATTACATTTTAAATATGCGAAGTGGATAGGAGAACTGCAGCT 50
|||||
533 GlylleHislleuLyslleMetTrpLysTrpIleGlyGluLeuGlnLe 549

51 GTCATATAGCC 60
|||||
549 uSerlleAla 552

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAG75614

seq_documentation_block:

ID AAG75614 standard; Protein: 869 AA.

AC AAG75614;

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO:6378.

DE Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.

OS Homo sapiens.

PN WO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

XX 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI; 2001-235357/24.

XX N-PSDB; AAH35019.

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -

XX Claim 11; Page 7851-7854; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAH77789 represent sequences used in the exemplification of the
XX present invention.

XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 869 AA;

alignment_scores:

Quality: 107.00

Length: 20

Ratio: 5.350 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-17 x AAG75614

Align seg 1/1 to: AAG75614 from: 1 to: 869

1 GGCATTACATTTTAAATATGCGAAGTGGATAGGAGAACTGCAGCT 50
|||||
850 GlylleHislleuLyslleMetTrpLysTrpIleGlyGluLeuGlnLe 866

51 GTCATATAGCC 60
|||||

866 uSerlleAla 869

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74824

seq_documentation_block:

ID AAB74824 standard; Protein: 914 AA.

AC AAB74824;

XX 13-JUN-2001 (first entry)

XX Human ICACC-1 protein sequence.

XX ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
XX interleukin 9 induced calcium activated chloride channel; IL-9;
XX calcium activated chloride channel; anti-allergic; anti-asthmatic;
XX anti-inflammatory; immunomodulatory; cystic fibrosis;
XX inflammatory bowel disease; autoimmune disease.

XX Homo sapiens.

XX WO9944620-A1.

XX 10-SEP-1999.

XX 03-MAR-1999; 99WO-US04703.

XX 03-MAR-1998; 98US-0076815.

XX (MAGA-) MAGALININ PHARM INC.

XX Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;

XX PI Nicolalides NC, Zhou Y, Dong Q;

XX WPI; 1999-550979/46.

XX N-PSDB; AAF81927.

XX New nucleic acid encoding calcium activated chloride channel, used to
XX identify, e.g. specific modulators for treating atopic allergy -

XX Claim 11; Fig 4B; 75pp; English.

XX The present sequence represents the human interleukin 9 (IL-9) induced
XX calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
XX have anti-allergic, anti-asthmatic, anti-inflammatory and
XX immunomodulatory activities. Compounds (A) that downregulate ICACC are
XX used to alleviate asthma (or more generally atopic allergy), while those
XX (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
XX inflammatory bowel disease (IBD) (or other autoimmune diseases).
XX Measuring levels of ICACC-1 can be used in the diagnosis of asthma
XX (increased levels) or IBD (reduced levels), also for monitoring
XX treatment of these conditions. The ICACC proteins can be used:
XX (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
XX reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
XX (A) to disrupt bonding between ICACC and its ligands; and (iii) to
XX identify modulators and binding partners. ICACC polynucleotides can be
XX used to generate transgenic animals or recombinant cells, used to screen
XX for antagonists, also as a source of therapeutic antisense agents or
XX diagnostic probes (for quantifying mRNA expression, e.g. for

CC identification of modulators).

XX Sequence 914 AA;

alignment_scores:
Quality: 107.00 Length: 20
Ratio: 5.350 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-17 x AAB74824 ..

Align seg 1/1 to: AAB74824 from: 1 to: 914

1 GGCATTACATTTTAAATATGCGAAGTGGATAGGAGAACTGCAGCT 50
|||||
895 GlyIleHisIleLeuLysIleMetTrpLysTrpIleGlyGluLeuGlnLe 911

51 GTCAATAGCC 60

|||||

911 uSerIleAla 914

seq_name: /SIDS2/gcgdata/geneseq/AA2001.DAT:AA24514

seq_documentation_block:

ID_AAM24514 standard; Protein: 914 AA.

XX AC AAM24514;

XX AC AAM24514;

XX AC AAM24514;

DT 12-OCT-2001 (first entry)

XX C902P predicted amino acid sequence.

DE Human; immunotherapy; diagnosis; colon cancer; colon tumour;

KW immunogenic; gene therapy; vaccine; colonic cancer.

KW Homo sapiens.

XX WO200149716-A2.

XX 12-JUL-2001.

XX 29-DEC-2000; 2000WO-US35596.

XX 30-DEC-1999; 99US-0476296.

XX 10-JAN-2000; 2000US-0480321.

XX 15-FEB-2000; 2000US-0504629.

XX 06-MAR-2000; 2000US-0519444.

XX 19-MAY-2000; 2000US-0575251.

XX 29-JUN-2000; 2000US-0609448.

XX 28-AUG-2000; 2000US-0649811.

XX (CORI-) CORIXA CORP.

XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;

PI King GE, Wang T, Jiang Y;

XX WPI; 2001-441847/47.

XX Colon tumor associated proteins and nucleic acids useful for the

PT prevention, diagnosis and treatment of colonic cancer -

XX Claim 2; Page 440-443; 472pp; English.

XX The present invention describes colon tumour associated proteins (I) and

CC the polynucleotides (II) that encode them. (I) have cytostatic activity.

CC (I) and (II) may be used in gene therapy and vaccine production. (I) and

CC (II) may be used in the prevention, diagnosis and treatment of diseases

CC associated with inappropriate colon tumour associated protein (TCAP)

CC expression, such as colonic cancer. For example, (I) and (II) may be

CC used to treat disorders associated with decreased expression by

CC rectifying mutations or deletions in a patient's genome that affect the

CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host
CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used-as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.

XX Sequence 914 AA;

alignment_scores:
Quality: 107.00 Length: 20
Ratio: 5.350 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-17 x AAM24514 ..

Align seg 1/1 to: AAM24514 from: 1 to: 914

1 GGCATTACATTTTAAATATGCGAAGTGGATAGGAGAACTGCAGCT 50
|||||
895 GlyIleHisIleLeuLysIleMetTrpLysTrpIleGlyGluLeuGlnLe 911

51 GTCAATAGCC 60

|||||

911 uSerIleAla 914

seq_name: /SIDS2/gcgdata/geneseq/AA2001.DAT:AAB73716

seq_documentation_block:

ID_AAB73716 standard; Protein: 914 AA.

XX AC AAB73716;

XX AC AAB73716;

XX AC AAB73716;

DT 11-SEP-2001 (first entry)

XX Human CLCA1 protein, SEQ ID NO:2.

DE Human CLCA1 protein, SEQ ID NO:2.

XX Human CLCA1; goblet cell; mouse Gob-5 orthologue; drug screening;

KW expression inhibition; antisense therapy; gene therapy;

KW chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic.

XX Homo sapiens.

XX Key Location/Qualifiers

PH Key Location/Qualifiers

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

FT Misc-difference 152

```
DR N-PSDB; AAH46102, AAH46124.
XX
PT New antisense nucleotide, useful for treatment and prevention of
XX bronchial asthma and chronic obstructive pulmonary disease
XX
PS Claim 2; Page 76-80; 104pp; Japanese.
XX
CC The invention relates to an antisense nucleotide targetted to the mouse
CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
CC the CLCA1 gene (coding sequence shown in AAH46102). The invention also
CC relates to an antibody specific for the Gob-5 protein, medical and
CC diagnostic compositions containing the antisense nucleotide or the
CC antibody, and methods and kits for screening for compounds which inhibit
CC the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.
CC The antisense oligonucleotides and antibody are therefore useful for the
CC treatment and prevention of bronchial asthma and chronic obstructive
XX pulmonary disease. The present sequence represents human CLCA1 protein.
XX
SQ Sequence 914 AA:

alignment_scores:
    Quality: 107.00      Length: 20
    Ratio: 5.350        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-049-696-17 x AAB73716 ..
Align seg 1/1 to: AAB73716 from: 1 to: 914

1 GGCATTACATTTTAAATAATATGTGGAAGTGGATGAGGAGCACTGCAGCT 50
|||||
895 GlyIleHisIleuLysIleMetTrpLysTrpIleGlyGluLeuGlnLe 911

51 GTCATATAGCC 60
|||||
911 uSerIleAla 914

seq_name: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT: AAB74733
seq_documentation_block:
ID AAB74733 standard; Protein; 914 AA.
AC AAB74733;
DE 12-JUN-2001 (first entry)
XX Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.
XX
XX Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
KW antialzheimer; antiparkinsonian; antimicrobial; vulnary; gene therapy;
KW immune disorder; hyperproliferative disorder; cardiovascular disease;
KW cancer; angiogenic disorder; neurological disorder; infectious disease;
KW wound healing; regeneration; chemotaxis; chromosome 1.
XX
OS Homo sapiens.
XX
XX WO200112775-A2.
XX
XX 22-FEB-2001.
PD
XX 16-AUG-2000; 2000WO-US22325.
XX
XX 17-AUG-1999; 99US-0149182.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ni J, Florence KA, Fiscella M, Wei P, Baker KP;
PI Birse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;
```

```
XX
DR N-PSDB; 2001-147550/15.
XX N-PSDB; AAF81787.
XX
PT Nucleic acids encoding 25 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
XX Claim 11; Page 459-460; 485pp; English.
XX
XX AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733
XX to AAB74772. Human secreted proteins can have activities based on the
XX tissues and cells they are expressed in. Example of activities include:
XX immunomodulatory; antisclerotic; dermatological; immunosuppressive;
XX antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;
XX vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic;
XX anticonvulsant; antialzheimer; antiparkinsonian; antimicrobial; and
XX vulnary. Human secreted proteins can be used in gene therapy and
XX vaccine. Human secreted protein nucleotide sequences (NAM1) and proteins
XX (PEP1) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate polypeptide expression. For example, NAM1
XX and PEP1 may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patients genome
XX that affect the activity of proteins by expressing inactive proteins or
XX to supplement the patients own production of polypeptides. Disorders that
XX may be prevented, diagnosed and/or treated include immune disorders,
XX hyperproliferative disorders (e.g. cancers), cardiovascular diseases,
XX angiogenic disorders, neurological disorders, infectious diseases and/or
XX for promoting wound healing, regeneration and /or chemotaxis. AAF81778 to
XX AAF81786 and AAB74732 represent sequences used in the exemplification of
XX the present invention.
XX
SQ Sequence 914 AA:

alignment_scores:
    Quality: 107.00      Length: 20
    Ratio: 5.350        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-049-696-17 x AAB74733 ..
Align seg 1/1 to: AAB74733 from: 1 to: 914

1 GGCATTACATTTTAAATAATATGTGGAAGTGGATGAGGAGCACTGCAGCT 50
.: |||||
895 GlyIleHisIleuLysIleMetTrpLysTrpIleGlyGluLeuGlnLe 911

51 GTCATATAGCC 60
.: |||||
911 uSerIleAla 914

seq_name: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT: AAG75474
seq_documentation_block:
ID AAG75474 standard; Protein; 925 AA.
XX
XX AAG75474;
XX
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen protein SEQ ID NO:6238.
DE
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 1.
XX
XX Homo sapiens.
XX
XX WO200122920-A2.
XX
XX 05-APR-2001.
XX
```

PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI; 2001-235357/24.
DR N-PSDB; AAH34879.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 11; Page 7686-7690; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 925 AA;
SQ

alignment_scores:
Quality: 107.00 Length: 20
Ratio: 5.350 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-17 x AAG75474 ..

Align seg 1/1 to: AAG75474 from: 1 to: 925

1 GGCATTACATTTTAAAAAATTATGTGGAGTGGATAGGAGAACTGCAGCT 50
|||||
906 GlyIleHisIleLeuLysileMetTrpLysTrpIleGlyGluLeuGlnLe 922

51 GTCATAGCC 60
|||||
922 useIleala 925

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.AAB74822

seq_documentation_block:
ID AAB74822 standard; Protein; 913 AA.
XX
AC AAB74822;
XX
XX 13-JUN-2001 (first entry)
DT
XX Murine ICACC-1 protein sequence.
DE
XX ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;

KW inflammatory bowel disease; autoimmune disease.
XX
OS Mus sp.
XX
XX WO944620-A1.
XX
XX 10-SEP-1999.
XX
XX 03-MAR-1999; 99WO-US04703.
XX
XX 03-MAR-1998; 98US-0076815.
XX
XX (MAGA-) MAGAININ PHARM INC.
XX
XX Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;
PI Nicolaides NC, Zhou Y, Dong Q;
XX
XX WPI; 1999-550979/46.
DR N-PSDB; AAF81925.
XX
XX New nucleic acid encoding calcium activated chloride channel, used to
PT identify, e.g. specific modulators for treating atopic allergy -
XX
XX Claim 12; Fig 2; 75pp; English.
XX
XX The present sequence represents the murine interleukin 9 (IL-9) induced
CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
CC have anti-allergic, anti-asthmatic, anti-inflammatory and
CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
CC used to alleviate asthma (or more generally atopic allergy), while those
CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
CC (increased levels) or IBD (reduced levels), also for monitoring
CC treatment of these conditions. The ICACC proteins can be used:
CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
CC reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
CC identify modulators and binding partners. ICACC polynucleotides can be
CC used to generate transgenic animals or recombinant cells, used to screen
CC for antagonists, also as a source of therapeutic antisense agents or
CC diagnostic probes (for quantifying mRNA expression, e.g. for
CC identification of modulators).
XX
XX Sequence 913 AA;
SQ

alignment_scores:
Quality: 90.00 Length: 19
Ratio: 4.737 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 68.421

alignment_block:
US-09-049-696-17 x AAB74822 ..

Align seg 1/1 to: AAB74822 from: 1 to: 913

1 GGCATTACATTTTAAAAAATTATGTGGAGTGGATAGGAGAACTGCAGCT 50
|||||
692 GlyIleHisIleLeuLysileMetTrpLysTrpLeuGlyGluMetGlnVa 908

51 GTCATA 57
|||||
908 lThrLeu 910

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.AAB73715

seq_documentation_block:
ID AAB73715 standard; Protein; 913 AA.
XX
AC AAB73715;
XX
XX 11-SEP-2001 (first entry)
DT

This Page Blank (uspto)

OM of: US-09-049-696-16 to: SPTREMBL_17:* out_format : pfs

Date: Mar 30, 2002 2:47 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q/cqn2_1/USPTO_spool/US09049696/runat_28032002_145238_2053/app_query.fasta_1.12579
-DB=SPTREMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-XGAPOP=6.000 -XGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -ECGN1_1_705 -NCPU=6 -LCPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-16
Query length: 242
Database: SPTREMBL_17:*
Database sequences: 473505
Database length: 146272329
Search time (sec): 805.760000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
sp_human:O95151	+	278.00	650.01	6.8e-29	914	I 095151 homo sapiens (human). ca
sp_human:O9URC6	+	278.00	650.01	6.8e-29	914	I 09urc6 homo sapiens (human). ca
sp_human:O9UNF6	+	278.00	650.01	6.8e-29	914	I 09unf6 homo sapiens (human). ca
sp_mammal:O9TUB5	+	212.50	490.80	5.0e-20	917	I 09tub5 sus scrofa (pig). epithe
sp_rodent:O98826	+	178.50	410.63	1.5e-15	913	I 098826 mus musculus (mouse). g
sp_rodent:O9D726	+	179.50	410.63	1.5e-15	913	I 09d726 mus musculus (mouse). ch
sp_invertebrate:O9W509	+	73.00	154.61	0.3910	626	I 09w509 drosophila melanogaste
sp_rodent:O9S820	+	72.00	152.12	0.5339	631	I 09s820 rattus norvegicus (rat).
sp_invertebrate:O9W4A2	+	71.00	149.78	0.7296	623	I 09w4a2 drosophila melanogaste
sp_invertebrate:O27423	+	68.50	151.05	1.67	232	I 027423 drosophila virilis (fr
sp_invertebrate:O9UR81	+	68.00	144.01	1.88	508	I 09ur81 drosophila melanogaste
sp_invertebrate:O9VXX5	+	67.50	141.09	2.17	639	I 09vxx5 drosophila melanogaste
sp_plant:O9LJL2	+	67.00	137.83	2.50	841	I 09ljl2 arabidopsis thaliana (md
sp_virus:O9DWM0	+	66.00	147.83	3.70	158	I 09dwm0 pluttella xylostella gran
sp_human:O9HAL3	+	66.00	141.11	3.54	390	I 09hal3 homo sapiens (human). cd
sp_human:O9UG75	+	66.00	139.23	3.50	502	I 09ug75 homo sapiens (human). hy
sp_human:O9J602	+	66.00	135.27	3.42	856	I 09j602 homo sapiens (human). el
sp_human:O9BUJ2	+	66.00	135.27	3.42	856	I 09bu2 homo sapiens (human). el
sp_invertebrate:O24459	+	66.00	135.26	3.42	857	I 024459 drosophila melanogaste
sp_invertebrate:O9V8C2	+	66.00	133.81	3.38	1042	I 09v8c2 drosophila melanogaste
sp_invertebrate:O9VZC2	+	65.50	138.90	4.11	446	I 09vzc2 drosophila melanogaste
sp_invertebrate:O23851	+	65.50	132.03	3.94	1123	I 023851 dictyostelium discoi
sp_bacteria:O9AOG1	+	64.50	142.54	5.84	197	I 09aog1 caldicellulosiruptor sp.
sp_bacteria:O9AQ66	+	64.50	142.36	5.83	202	I 09aq66 caldicellulosiruptor sp.
sp_rodent:O90749	+	64.50	136.52	5.62	443	I 090749 mus musculus (mouse). p
sp_rodent:O90735	+	64.50	136.52	5.62	443	I 090735 mus musculus (mouse). sh
sp_rodent:O9W333	+	64.50	136.52	5.62	443	I 09w333 mus musculus (mouse). sh
sp_plant:O01927	+	64.00	138.18	6.69	301	I 001927 phytophthora infestans
sp_vertebrate:O91767	+	64.00	137.67	6.67	322	I 091767 xenopus laevis (afri
sp_plant:O01916	+	64.00	137.12	6.84	347	I 001916 phytophthora infestans
sp_virus:O9WRM2	+	64.00	135.22	6.57	448	I 09wr2 macaca mulatta rhadinovi
sp_virus:O9J2H0	+	64.00	135.22	6.57	448	I 09j2h0 macaca mulatta rhadinovi
sp_plant:O65672	+	64.00	132.18	6.44	674	I 065672 arabidopsis thaliana (md
sp_invertebrate:O9W384	+	64.00	131.40	6.41	749	I 09w384 drosophila melanogaste
sp_invertebrate:O9BIW6	+	64.00	131.40	6.41	749	I 09biw6 drosophila melanogaste
sp_plant:O9C660	+	64.00	131.29	6.41	760	I 09c660 arabidopsis thaliana (md
sp_plant:O9LNT8	+	64.00	128.30	6.29	1137	I 09lnt8 arabidopsis thaliana (md
sp_invertebrate:O45328	+	63.50	133.22	7.64	498	I 045328 caenorhabditis elegans
sp_bacteria:O9K5H9	+	63.00	143.65	9.60	104	I 09k5h9 bordetella bronchiseptid
sp_bacteria:O9K5H7	+	63.00	143.57	9.60	105	I 09k5h7 bordetella bronchiseptid

sp_bacteria:O9K5G5 + 63.00 143.57 9.60 105 I 09k5g5 bordetella bronchisept
sp_bacteria:O9K5H5 + 63.00 143.43 9.59 107 I 09k5h5 bordetella bronchisept
sp_bacteria:O9K5H1 + 63.00 143.30 9.58 109 I 09k5h1 bordetella bronchisept
sp_bacteria:O9K5G9 + 63.00 143.16 9.57 111 I 09k5g9 bordetella bronchisept
sp_bacteria:O9K5G7 + 63.00 143.16 9.57 111 I 09k5g7 bordetella bronchisept

seq_name: sp_human:O95151

seq_documentation_block:

ID O95151 PRELIMINARY; PRT; 914 AA.
AC O95151;
DT 01-MAY-1999 (TREMREL. 10, Created)
DT 01-MAY-1999 (TREMREL. 10, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCA1.
OS Hmo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Eble R.C., Ji H.L., Schreuer K.D., Fuller C.M.,
RA Pauli B.O.;
RT "Genomic cloning, molecular characterization, and functional analysis
of human HCLCA1, the first human member of the family of Ca2+-activated
Cl-channel proteins.";
RL Genomics 54:200-214(1998).
DR EMBL; AF039400; AAC95428.1; -;
DR InterPro; IPR000131; ATPase_gamma.
DR InlerPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS00234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 914 AA; 100174 MW; 7066440C46557FA3 CRC64;

alignment_scores:

Quality: 278.00 Length: 50
Ratio: 5.560 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-16 x O95151 ..
Align seg 1/1 to: O95151 from: 1 to: 914
.. 2 TTTATTCTCCACAGACTCCGCCAGACACCTAGTCTGTGATGAACGTC 51
|||||
865 PhelieProProGlnThrProProGluThrProSerProAspGluThrse 881
.. 52 TGCTCCCTGTGCTTAATATTCATATCAACAGCACCATCTCTGGCATTCACA 101
|||||
881 rAlaproCysProAsnIleHisIleAsnSerThIleProGlyIleHisI 898
.. 102 TTTTAAAAATATGTGGAAGTGGATAGGAGAACTGCAGCTGTCAATAGCC 151
|||||
898 leLeuLysIleMetTrpIysTrpIleGlyGluLeuGlnLeuSerIleala 914

seq_name: sp_human:O9UPC6

seq_documentation_block:

ID O9UPC6 PRELIMINARY; PRT; 914 AA.
AC O9UPC6;
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCA1.
OS Hmo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99047526; PubMed=9828122;
 RA Gruber A.D., Elble R.C., Ji H.L., Schreier K.D., Fuller C.M.,
 RA Pauli B.U.;
 RT "Genomic cloning, molecular characterization, and functional analysis
 RT of human CLCA1, the first human member of the family of Ca2+-activated
 RT Cl-channel proteins.";
 RL Genomics 54:200-214(1998).
 DR EMBL; AF039401; AAC95429.1; ..
 DR InterPro; IPR000131; ATPase_gamma.
 DR InterPro; IPR002035; vWFA.
 DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
 DR PROSITE; PS50234; vWFA; 1.
 DR SMART; SM00327; vWA; 1.
 SQ SEQUENCE 914 AA; 100160 MW; 7761362B66752AA3 CRC64;

alignment_scores:
 Quality: 278.00 Length: 50
 Ratio: 5.560 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-16 x Q9UPC6 ..
 Align seg 1/1 to: Q9UPC6 from: 1 to: 914

2 TTTATCTCTCCACAGACTCGCCGAGACACCTAGTCTGATGAACGTC 51
 |||||||
 865 PheileProGlnThrProGluThrProSerProAspGluThrSe 881
 52 TGCTCTCTGCTTAATTCATATCATACACGACCATTCCTGGCATTCACA 101
 |||||||
 881 rAlaProCysProAsnIleHisIleAsnSerThrIleProGlyIleHisI 898
 102 TTTTAAATATATGTGAAGTGGATAGGAGAACTGCAGCTGTCAATAGCC 151
 |||||||
 898 leLeuLysIleMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914

seq_name: sp_human:Q9UNF6

seq_documentation_block:

ID Q9UNF6 PRELIMINARY; PRT; 914 AA.
 AC Q9UNF6;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 1.
 GN CACCI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SMALL INTESTINE, AND COLON;
 RX MEDLINE=99364503; PubMed=10437792;
 RA Agnel M., Vermet T., Culoucou J.M.;
 RT "Identification of three novel members of the calcium-dependent
 RT chloride channel (CaCC) family predominantly expressed in the
 RT digestive tract and trachea.";
 RL FEBS Lett. 455:295-301(1999).
 DR EMBL; AF127036; AAD25487.1; ..
 DR InterPro; IPR000131; ATPase_gamma.
 DR InterPro; IPR002035; vWFA.
 DR SMART; SM00327; vWA; 1.
 DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
 DR PROSITE; PS50234; vWFA; 1.
 SQ SEQUENCE 914 AA; 100202 MW; B58B451D97887D6B CRC64;

alignment_scores:
 Quality: 278.00 Length: 50
 Ratio: 5.560 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-049-696-16 x Q9UNF6 ..
 Align seg 1/1 to: Q9UNF6 from: 1 to: 914
 2 TTTATCTCTCCACAGACTCGCCGAGACACCTAGTCTGATGAACGTC 51
 |||||||
 865 PheileProGlnThrProGluThrProSerProAspGluThrSe 881
 52 TGCTCTCTGCTTAATTCATATCATACACGACCATTCCTGGCATTCACA 101
 |||||||
 881 rAlaProCysProAsnIleHisIleAsnSerThrIleProGlyIleHisI 898
 102 TTTTAAATATATGTGAAGTGGATAGGAGAACTGCAGCTGTCAATAGCC 151
 |||||||
 898 leLeuLysIleMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914

seq_name: sp_mammal:Q9TUB5

seq_documentation_block:

ID Q9TUB5 PRELIMINARY; PRT; 917 AA.
 AC Q9TUB5;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE EPITHELIAL CHLORIDE CHANNEL PROTEIN.
 GN AECC.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=ILEAL MUCOSA;
 RA Gaspar K.J., Gabriel S.E., Racette K.J., Forsyth G.W.;
 RT "A cDNA involved in porcine exocrine chloride conductance.";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF095584; AAF00077.1; ..
 DR InterPro; IPR000131; ATPase_gamma.
 DR InterPro; IPR002035; vWFA.
 DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
 DR PROSITE; PS50234; vWFA; 1.
 DR SMART; SM00327; vWA; 1.
 SQ SEQUENCE 917 AA; 100735 MW; 649F41944F7EAD19 CRC64;

alignment_scores:
 Quality: 212.50 Length: 48
 Ratio: 4.830 Gaps: 1
 Percent Similarity: 91.667 Percent Identity: 79.167
 alignment_block:
 US-09-049-696-16 x Q9TUB5 ..
 Align seg 1/1 to: Q9TUB5 from: 1 to: 917
 9 CPTCCACAGACTCGCCGAGACACCTAGTCTGATGAACGTCCTGCTCC 57
 |||||||
 873 ProProGluThrProGluThrProAlaPro.....SerLeuPr 886
 58 TTGCTCTTAATTCATATCATACACGACCATTCCTGGCATTCACATTTAA 107
 |||||||
 886 oCysProGluIleGlnValAsnSerThrIleProGlyIleHisIleLeuL 903
 108 AAATATGCTGAAGTGGATAGGAGAACTGCAGCTGTCAATAGCC 151
 |||||||
 903 ysIleMetTrpLysTrpLeuGlyGluLeuGlnLeuSerIleAla 917

RC	STRAIN=C57BL/6J; TISSUE=STOMACH;	
RX	MEDLINE=21085660; PubMed=11217851;	
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,	RA Begos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D., Barrell B., RA Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E., Dreano S., RA Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D., Minana B.,

RA Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
RA Modolell J., Peter A., Schottler P., Werner M., Mourkioti F.,
RA Belinetti N., Dowe G., Schafer U., Jackle H., Bucheton A.,
RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
RA McWilliam P.J., Salles C., Tait E.A., Valenti P., Saunderson R.D.,
RA Glover D.M.;
RT "From sequence to chromosome: the tip of the X chromosome of D.
RT melanogaster";
RL Science 287:2220-2222(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doug L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laško P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
DR EMBL: AL009193; CAA15693.1; ..
DR EMBL: AE003423; AAF45749.2; ..
DR FlyBase: FBgn0023549; Mct1.
DR InterPro: IPR002897; MCT.
DR Pfam: PF01587; MCT; 1.
SQ SEQUENCE 626 AA; 67913 MW; 30FEFCF6B4E8B0B4 CRC64;

alignment_scores:
Quality: 73.00 Length: 67
Ratio: 1.780 Gaps: 2
Percent Similarity: 61.194 Percent Identity: 32.836

alignment_block:

US-09-049-696-16 x Q9W509

Align seg 1/1 to: Q9W509 from: 1 to: 626

8 CCTCCAGACTCCGCGACAGACACCTAGTCCTGATGAACGTCCTGCTCC 57
|||||
218 ProGlnThrProAsnThrProSerThrProLysLeuAlaLysLys 234

58 TTGTCCTTAATATTCATATCAACAGACACCATTCCTGGCATTACATTTAA 107
||| |||
234 sserAlaIleleAspGlnAsnThrThraAlaGluLeuGluProLeuL 251
|||
108 AAAT.....TATGCGGAAGTGGATAGGAGAACTGCAGCT 141
||
251 ysValLeuProLysAspGlnTyrlLeuGlnLeuProGlnArgThrAspThr 267
|||
142 GTCATAGCTAGGCTCAATTTTGTTCAGATAATAATAATCAATCATTC 191
|||
268 AlaValSerGlyGlyAla....LeuCysArgSerAsnSerValGlyHisAs 283
|||
192 A 192
283 n 283

seq_name: sp_rodent:Q9ESZ0

seq_documentation_block:

ID Q9ESZ0 PRELIMINARY; PRT; 631 AA.

AC Q9ESZ0;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE X-RAY REPAIR CROSS-COMPLEMENTING GROUP 1 PROTEIN XRCC1.

GN XRCC1.

CS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN;

RA Chen D., Cao G., Li M., Zhang Y., Glenn G., Chen J.;

RT "Cloning and characterization of rat Xccc-1, a DNA repair gene in

development and aging of rat.";

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF290895; AAG0212.1; ..

DR InterPro: IPR001357; BRCT.

DR InterPro: IPR002706; XRCC1_N.

DR Pfam: PF00533; BRCT; 2.

DR Pfam: PF01834; XRCC1_N; 1.

DR ProDom: PD023136; XRCC1_N; 1.

DR SMART: SM00292; BRCT; 2.

DR PROSITE: PS00172; BRCT; 2.

SQ SEQUENCE 631 AA; 68810 MW; 99C01B9A41F22BE4 CRC64;

alignment_scores:
Quality: 72.00 Length: 28
Ratio: 3.429 Gaps: 0
Percent Similarity: 75.000 Percent Identity: 46.429

alignment_block:

US-09-049-696-16 x Q9ESZ0

Align seg 1/1 to: Q9ESZ0 from: 1 to: 631

11 CCACAGACTCCGCGACAGACACCTAGTCCTGATGAACGTCCTGCTTG 60
|||||
444 ProSerSerProProArgProProThrProGluGluThrLysAlaProSe 460

61 TCCTAATATTCATATCAACAGACACCATTCCTGCGC 94
|||||

460 rProGlyProGlnAspAsnSerAspThrAspGly 471

seq_name: sp_invertebrate:Q9W4A2

seq_documentation_block:

ID Q9W4A2 PRELIMINARY; PRT; 623 AA.

AC Q9W4A2;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

[illegible]

Quality: 68.50 Length: 31
 Ratio: 3.114 Gaps: 2
 Percent Similarity: 70.968 Percent Identity: 48.387

alignment_block:

US-09-049-696-16 x Q27423 ..
 Align seg 1/1 to: Q27423 from: 1 to: 232
 8 CTTCCACAGACTCCGCCAGACGAG.....ACACCTAGTCTGATGAA...AC 48
 |||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 28 ProProGluThrThrProGluProThrThrProGluProCysGluThrTh 44
 49 CTTCTGCTCTTGTCTTAATATTCATATCAACAGACCACTTCCT 91
 |||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 44 rThrThrProCysProThrThrThrThrThrThrThrThrPro 58

seq_name: sp_invertebrate:Q9UB81

seq_documentation_block:

ID Q9UB81 PRELIMINARY; PRT; 508 AA.
 AC Q9UB81;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE MONOCARBOXYLATE TRANSPORTER 1 HOMOLOGUE.
 GN MCT1 OR EG:103B4.3 OR CG3456.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kovalenko T.A., Alatorsev V.E.;
 RT "dMct1, Drosophila gene having translation product similar to the
 RL monocarboxylate transporter 1 protein.";
 DR EMBL: AJ328706; CAB42050.1;
 DR FlyBase: FBgn0023549; MCT1.
 DR InterPro: IPR002897; MCT.
 DR Pfam: PF01587; MCT; 1.
 SQ SEQUENCE 508 AA; 54912 MW; C4322FA21B795090 CRC64;

alignment_scores:

Quality: 68.00 Length: 67
 Ratio: 1.659 Gaps: 2
 Percent Similarity: 61.194 Percent Identity: 31.343

alignment_block:

US-09-049-696-16 x Q9UB81 ..
 Align seg 1/1 to: Q9UB81 from: 1 to: 508
 8 CTTCCACAGACTCCGCCAGACACTAGTCTCTGATGAACGTCGTCTCC 57
 |||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 100 ProProGlnThrProProAsnThrProSerThrProLysLeuAlaLysLy 116
 58 TTGTCCTAATATTCATATCAACAGACCACTTCCTGGCATTCACATTTAA 107
 |||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 116 sSerAlaIleAspGlnAsnThrAlaAlaAlaGluLeuGluProLeuL 133
 108 AAAT.....TATGTGGAGTGGATAGGAGAACTGCAGCT 141
 |||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 133 ysValLeuProLysAspGlnThrLeuGlnLeuProGlnArgThrAspThr 149
 142 GPCAAATAGCTAGGCGCTGAATTTTGTTCAGATAAATAAATCAATC 191
 |||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 150 AlaValSerGlyGlyAla....LeuCysArgSerAsnSerValGlyHis 165
 192 A 192

165 n 165

seq_name: sp_invertebrate:Q9VXX5

seq_documentation_block:

ID Q9VXX5 PRELIMINARY; PRT; 639 AA.
 AC Q9VXX5;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE CGI5029 PROTEIN.
 GN CGI5029.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Bookova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Esler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Hariri N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Meunt S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.C.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 ER EMBL: AE003498; AAF48431.2; -;
 ER FlyBase: FBgn0030619; CGI5029.
 SQ SEQUENCE 639 AA; 67593 MW; 53BB977916B400C7 CRC64;

alignment_scores:

Quality: 67.50 Length: 36
 Ratio: 3.068 Gaps: 1
 Percent Similarity: 61.111 Percent Identity: 38.889

alignment_block:

US-09-049-696-16 x Q9VXX5 ..

Align seg 1/1 to: Q9VXX5 from: 1 to: 639

01-MAR-2001 (TREMBlé, 16, Created)

alignment_scores

alignment_scores

```
Quality: 66.00      Length: 35
Ratio: 3.300        Gaps: 1
Percent Similarity: 57.143  Percent Identity: 40.000

alignment_block:
US-09-049-696-16 x Q9HAL3  ..

Align seg 1/1 to: Q9HAL3 from: 1 to: 390

      8 CCTCCACAGACTCCGCCAGAGACACCTAGTCTGTGATGAACGCTGTGCTCC 57
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
227 ProProGlnProProGlnProProGlnProProProProSerTyrSerPr 243
58 TTGTCCTAAT.....ATTGATATCAACAGCACCATTTC 89
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
243 OAlaArgAsnProProGlyAlaSerThrTyrAsnLysAsnSerAsnIleP 260
90 CTGGC 94
   |||||
260 rogly 261
```

OM of: US-09-049-696-16 to: SwissProt_39.* out_format : pfs

Date: Mar 30, 2002 2:52 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=framet-n2p.model -DEV=xlp
-Q/cdn2_1/USPT0001/US09049696/runat_28032002_145238_2085/app_query.fasta_1.12579
-DB=SwissProt_39 -QFM=fastan -SUFFIX=isp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696.ecgnl_1.165 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-16
Query length: 242
Database: SwissProt_39.*
Database sequences: 100059
Database length: 36664827
Search time (sec): 306.030000

score.list:

Sequence	Strid Orig	ZScore	EScore	Len	Documentation
SwissProt_39:XRCL_MOUSE +	69.00	134.91	1.22	631	! Q60596 mus musculus (mouse).
SwissProt_39:YWS_WHEAT +	64.50	135.93	4.46	151	! P14723 triticum aestivum (wh
SwissProt_39:APG_ARATH +	64.00	125.09	5.07	534	! P40602 arabidopsis thaliana (
SwissProt_39:Y242_MYCPN +	63.00	121.57	6.72	632	! P75440 mycoplasma pertussis
SwissProt_39:PERT_BORPE +	63.00	118.76	6.69	910	! P14283 bordetella pertussis
SwissProt_39:PERT_BORBR +	63.00	118.75	6.69	911	! Q03035 bordetella bronchise
SwissProt_39:PERT_BORPA +	63.00	118.66	6.69	922	! P24328 bordetella parapertu
SwissProt_39:HCK_HUMAN +	62.50	121.87	7.77	526	! P08631 homo sapiens (human).
SwissProt_39:GUNG_DICDI +	62.50	119.61	7.74	705	! P22699 dictyostelium discoid
SwissProt_39:HXD1_XENLA +	62.00	125.08	9.02	300	! Q08820 xenopus laevis (afri
SwissProt_39:P54_ENTFC +	62.00	120.91	8.96	516	! P13692 enterococcus faecium (
SwissProt_39:N551_SOYBN +	62.00	116.34	8.90	933	! P07202 homo sapiens (human).
SwissProt_39:YV53_ROTBR +	61.00	128.90	12.11	137	! Q05544 glycine max (soybean).
SwissProt_39:ANDR_RAT +	61.00	119.07	11.93	491	! P12475 bovine rotavirus (str
SwissProt_39:ANDR_MOUSE +	61.00	114.41	11.84	899	! P19091 mus musculus (mouse).
SwissProt_39:PAR3_HUMAN +	61.00	114.38	11.84	902	! P15207 rattus norvegicus (rat
SwissProt_39:MUC2_HUMAN +	60.00	118.94	15.91	374	! Q00254 homo sapiens (human).
SwissProt_39:TEST_MOUSE +	59.50	97.59	17.77	5179	! Q02817 homo sapiens (human).
SwissProt_39:S3A2_HUMAN +	59.00	117.82	21.20	324	! Q9Jhj7 mus musculus (mouse).
SwissProt_39:NDVB_RHIME +	59.00	115.06	21.10	464	! Q15428 homo sapiens (human).
SwissProt_39:FXJ2_MOUSE +	59.00	114.88	21.10	475	! Q62203 mus musculus (mouse).
SwissProt_39:SYL_AQUAE +	58.50	112.43	24.28	2870	! P20471 rhizobium meliloti (
SwissProt_39:MKK2_MOUSE +	58.00	108.38	24.12	956	! Q6651 aquifex aeolicus. isol
SwissProt_39:APG_BRANA +	58.00	113.19	28.08	443	! P02831 mus musculus (mouse).
SwissProt_39:WCAM_ECOLI +	58.00	112.83	28.07	464	! P71244 escherichia coli. col
SwissProt_39:SPG7_DICDI +	58.00	112.15	28.04	507	! P49993 mycobacterium tubercu
SwissProt_39:PTGA_MYCGE +	58.00	107.66	27.90	532	! P22698 dictyostelium discoid
SwissProt_39:YME4_YEAST +	58.00	104.74	27.71	1328	! Q04715 mycoplasma genitalium
SwissProt_39:YMU0_YEAST +	58.00	104.74	27.71	1328	! Q04711 saccharomyces cerevi
SwissProt_39:UTX_MOUSE +	58.00	104.71	27.70	1333	! Q0546 mus musculus (mouse).
SwissProt_39:UTX_HUMAN +	58.00	104.32	27.69	1401	! O15550 homo sapiens (human).
SwissProt_39:BP11_YEAST +	58.00	103.50	27.65	1559	! P14772 saccharomyces cerevi
SwissProt_39:YJ27_YEAST +	58.00	102.59	27.61	1755	! P47098 saccharomyces cerevi
SwissProt_39:YJ29_YEAST +	58.00	102.59	27.61	1755	! P47100 saccharomyces cerevi

SwissProt_39:Y066_NPVLD + 57.50 113.49 32.46 369 ! P30325 lymantia dispar m
SwissProt_39:CGB3_CAEEL + 57.50 113.16 32.44 385 ! Q10654 caenorhabditis ele
SwissProt_39:FXJ2_HUMAN + 57.50 110.08 32.28 574 ! Q9p0k8 homo sapiens (huma
SwissProt_39:KKIT_FELCA + 57.50 105.98 32.07 978 ! Q28889 felis silvestris c
SwissProt_39:MAN2_MOUSE + 57.50 104.73 32.01 1150 ! P27046 mus musculus (mou

seq_name: SwissProt_39:XRCL_MOUSE

seq_documentation_block:

ID XRCL_MOUSE STANDARD; PRT; 631 AA.
AC Q60596;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DI 15-JUL-1999 (Rel. 38, Last annotation update)
DE DNA-REPAIR PROTEIN XRCC1.
GN XRCC1.
OS MUS musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
FN [!]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95048367; PubMed=7959765;
RA Brockman K.W., Tebbis R.S., Allen S.A., Tucker J.D., Swiger R.R.,
RA Lamerdin J.E., Carrano A.V., Thompson L.H.;
RT "Isolation and characterization of mouse Xrcc-1, a DNA repair gene
affecting ligation";
RL Genomics 22:180-188(1994).
CC -! FUNCTION: CORRECTS DEFECTIVE DNA STRAND-BREAK REPAIR AND SISTER
CC -! CHROMATID EXCHANGE FOLLOWING TREATMENT WITH IONIZING RADIATION
CC -! AND ALKYLATING AGENTS.
CC -! SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -! SIMILARITY: SOME, TO S.POMBE RAD4/CUT5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EXSL: U02887; AAA93115.1;
CC MGD; MGI:99137; XRCC1.
CC InterPro: IPR001357; BRCT.
CC InterPro: IPR002706; XRCC1_N.
CC Pfam: PF00533; BRCT; 2.
CC Pfam: PF01834; XRCC1_N; 1.
CC ProDom: PD023136; XRCC1_N; 1.
CC SMART: SM00292; BRCT; 2.
CC PROSITE: PS0172; BRCT; 2.
CC DNA REPAIR; Nuclear protein.
SQ SEQUENCE 631 AA; 69003 MW; 7F1868BFBEB2A3C68 CRC64;

alignment_scores:
Quality: 69.00 Length: 28
Ratio: 3.450 Gaps: 0
Percent Similarity: 71.429 Percent Identity: 46.429
alignment_block:
US-09-049-696-16 x XRCL_MOUSE ..
Align seg 1/1 to: XRCL_MOUSE from: 1 to: 631
11 CCAGACTCGCCAGACACCTAGTCCGATCAACGTCCTGCTG 60
444 ProSerProProAatGProProThProLysGluThrLysAlaProSe 460
61 TCCTTAATATTCATATCAACAGCACCATCCCGGC 94
460 rProGlyProGlnAspAsnSerAspThrGluGly 471

OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 CC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2104;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 29342 / M129;
 RX MEDLINE=97105885; PubMed=8948633;
 RA Hammelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
 RA Herrmann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 RT pneumoniae";
 RL Nucleic Acids Res. 24:4420-4449(1996).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE000049; AAB96146.1;
 DR Hypothetical protein; Transmembrane; Complete proteome.
 KW TRANSMEM 255 275 POTENTIAL.
 FT TRANSMEM 506 526 POTENTIAL.
 FT TRANSMEM 566 586 POTENTIAL.
 FT TRANSMEM 603 623 POTENTIAL.
 SQ SEQUENCE 632 AA; 74335 MW; 344D6771FCDAA44A6 CRC64;

alignment_scores:
 Quality: 63.00 Length: 77
 Ratio: 1.750 Gaps: 4
 Percent Similarity: 46.753 Percent Identity: 27.273

alignment_block:
 US-09-049-696-16 x Y242_MYCPN ..
 Align seg 1/1 to: Y242_MYCPN from: 1 to: 632

```

91 TGGCATTTCACATTTTAAATAATTATGTGAA.....GTGGATAG 128
   ||| :|||:||||| :|||: |||
402 TrpSerAsnHisPheValAspTrpGluPheLysAspLeuIleGluI 418
   : ||| ||| |||: |||
129 GAGAACTGCAGCTGTCAATAGCCTAGGGCTG..... 159
418 nIleThrAspAlaGluAsnGlySerAspLeuTyrGlyPheGluLysAspL 435
160 .....AATTTTGTGCAGATAATAA.....ATA 183
435 euAspGluSerIleCysGlnValAsnLysLysTyrLeuThrPheIleSer 451
184 AATCATTTCATCTTTTTCATTATAAA..... 213
452 SerAspSerSerPheLeuIleIleLysAsnAspGlnThrLysValII 468
214 .....TTTCTAAATGATTTT 231
468 eSerAsnTyrValTrpAlaGlnLeuTyrPhe 478

```

seq_name: SwissProt_39:PERT_BORPE

seq_documentation_block:
 ID PERT_BORPE STANDARD; PRT; 910 AA.
 AC P14283;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE PERTACTIN PRECURSOR (OUTER MEMBRANE PROTEIN P.69) (P.93).
 GN PRN OR OMP69A.
 OS Bordetella pertussis.

OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 CC Bordetella.
 OX NCBI_TaxID=520;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-ISOLATE CN2992;
 RX MEDLINE=89464462; PubMed=2542937;
 RA Charles I.G., Dougan G., Pickard D., Chatfield S., Smith M.,
 RA Novotny P., Morrissey P., Fairweather N.F.;
 RT "Molecular cloning and characterization of protective outer membrane
 RT protein P.69 from Bordetella pertussis";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:3554-3558(1989).
 RN [2]
 RP REVISIONS TO 264 AND 332.
 RX MEDLINE=92407514; PubMed=1527510;
 RA Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;
 RT "Cloning, nucleotide sequence and heterologous expression of the
 RT protective outer-membrane protein P.68 pertactin from Bordetella
 RT bronchiseptica";
 RL J. Gen. Microbiol. 138:1697-1705(1992).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE=96196517; PubMed=8609998;
 RA Emsley P., Charles I.G., Fairweather N.F., Isaacs N.W.;
 RT "Structure of Bordetella pertussis virulence factor P.69 pertactin";
 RL Nature 381:90-92(1996).
 CC -1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
 CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
 CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
 CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
 CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
 CC CONCENTRATION.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; J04560; AAA22980.1; ALT_SEQ.
 DR FIR; A32560; A32560.
 KW Outer membrane; Signal; Virulence; Repeat.
 FT SIGNAL 1 34
 FT CHAIN 35 910 P.93.
 FT CHAIN 35 711 PERTACTIN (P.69).
 FT PROPEP 712 910 POTENTIAL.
 FT SITE 260 262
 FT
 FT DOMAIN 266 290
 FT REPEAT 266 270 1.
 FT REPEAT 271 275 2.
 FT REPEAT 276 280 3.
 FT REPEAT 281 285 4 (APPROXIMATE).
 FT REPEAT 286 290 5 (APPROXIMATE).
 FT DOMAIN 579 593 5 X 3 AA TANDEM REPEATS OF P-Q-P.
 SQ SEQUENCE 910 AA; 93452 MW; A169871E20A2E7DB CRC64;

alignment_scores:
 Quality: 63.00 Length: 59
 Ratio: 1.969 Gaps: 4
 Percent Similarity: 54.237 Percent Identity: 32.203
 alignment_block:
 US-09-049-696-16 x PERT_BORPE ..
 Align seg 1/1 to: PERT_BORPE from: 1 to: 910

MoJ.	Cell. Biol.	7:2276-2285(1987).
RL	[3]	
RN		
RP		SEQUENCE OF 1-22 FROM N.A., AND ALTERNATIVE INITIATION.
RX		MEDLINE=91342636; PubMed=1875927;
RY		Lock P., Ralph S., Stanley E., Boulet I., Ramsay R., Dunn A.R.;
RA		"Two isoforms of murine hck, generated by utilization of alternative
RT		translational initiation codons, exhibit different patterns of
RT		subcellular localization.";
MoJ.	Cell. Biol.	11:4363-4370(1991).
RL	[4]	
RN		
RP		X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 78-526.
RX		MEDLINE=97177106; PubMed=9024658;
RY		Siéheri F., Moarefi I., Kuriyan J.;
RA		"Crystal structure of the Src family tyrosine kinase Hck.";
RT		Nature 385:602-609(1997).
RL	[5]	
RN		

RA Horita D.A., Baldissari D.M., Zhang W., Altieri A.S., Smithgall T.E.,
RA Gmeiner W.H., Byrd R.A.;
RT "Solution structure of the human Hck SH3 domain and identification of
RT its ligand binding site";

U.S. MOL. BIOL. 278:233-263 (1998).
PN [6]

STRUCTURE BY NMR OF 139-245.

RX MEJDLNE=9/263487; PubMed=9109402;
RA Zhang W, Smithbaall T F, Gmeiner W H.

RT sequential assignment and secondary structure determination for the src homolog 2 domain of hematopoietic cellular kinase".

RL FEBS Lett. 406:131-135(1997).
EN 174

RP X-RAY CRYSTALLOGRAPHY (2.6 ÅNGSTROMS) OF 81-137.
 BY MTCY TYNB-00452315. PUBV03-0270242.

RA Aröld S., O'Brien R., Franken P., Strub M.P., Hoh F., Dumas C.,

RT "R" loop flexibility enhances the specificity of Src family SH3

RL Biochemistry 37:14683-14691(1998).

RECEPTOR TO THE ACTIVATION OF THE RESPIRATORY BURST. MAY ALSO

CONTRIBUTE TO NEUTROPHIL MIGRATION AND MAY REGULATE THE DEGRANULATION PROCESS OF NEUTROPHILS.

1. CATALYTIC ACTIVITY: $\text{ATP} + \text{A PROTEIN TYROSINE} = \text{ADP} +$
2. PROTEIN TYROSINE PHOSPHATE.

-1:- SUBCELLULAR LOCATION: P60-HCK AND P59-HCK ARE ASSOCIATED WITH PROTEIN TYROSINE PHOSPHATASE.

MEMBRANES: P80-HCK IS ALSO CYTOPLASMIC (BY SIMILARITY).
ALTERNATIVE PRODUCTS: THE P60-HCK AND P59-HCK ARE PRODUCED BY THE

USE OF ALTERNATIVE INITIATION SITES.

CC-100: TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN CELLS OF THE MYELOID AND B-LYMPHOID LINEAGES.

CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.

CC 717 SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC-1: SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC

CC : DOMAIN. BELONGS TO THE SRC SUBFAMILY.

33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552

THIS SWISS-PROT entity is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the European Bioinformatics Institute.

between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC-0 or send an email to license@isb-sib.ch).

.....

DR EMBL; MI6591; AAA52643.1; -.
DR EMBL; MI6592; AAA52644.1; -.

DR EMBL; M16392; AAA32644.1; -.
DR DDBJ; M27813; TAYHHC

DRK: AZ7612, IVHHC:
PDH: 2HCK. 20-AUG-97

LEE, ZHCK, 29 AUG 97;
DR PD8: 3HCK: 15-OCT-97.

DR PDB: 4HCK; 17-JUN-98.

DR PDB; 5HCK; 17-JUN-98.

DR PDB; 1AD5; 15-MAY-97.

DR PDH; 1BUL; 11-NOV-98.

DR. MIN; 142370; -.

DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR000980; SH2.
 DR InterPro; IPR001452; SH3.
 DR InterPro; IPR001245; Tyr_kin.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00017; SH2; 1.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00109; TYRKINASE.
 DR PRINTS; PR00401; SH2DOMAIN.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR SMART; SM00252; SH2; 1.
 DR SMART; SM00326; SH3; 1.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS50001; SH2; 1.
 DR PROSITE; PS50002; SH3; 1.
 DR Transferase; Tyrosine-protein kinase; Phosphorylation; ATP-binding;
 KW Myristate; SH2 domain; SH3 domain; Alternative initiation;
 KW 3D-structure.
 FT CHAIN 1 526 TYROSINE-PROTEIN KINASE P60-HCK.
 FT INIT 22 526 TYROSINE-PROTEIN KINASE P59-HCK.
 FT INIT_MET 22 22 FOR P59-HCK.
 FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
 FT LIPID 23 23 MYRISTATE (BY SIMILARITY).
 FT DOMAIN 78 138 SH3.
 FT DOMAIN 144 241 SH2.
 FT DOMAIN 262 515 PROTEIN KINASE.
 FT NP_BIND 268 276 ATP.
 FT BINDING 290 290 ATP.
 FT ACT_SITE 381 381
 FT MOD_RES 411 411
 FT CONFLICT 24 24 S -> C (IN REF. 2; AAA52644).
 SQ SEQUENCE 526 AA; 59583 MW; 347E877A0A6412B3 CRC64;

alignment_scores:
 Quality: 62.50 Length: 29
 Ratio: 3.289 Gaps: 1
 Percent Similarity: 65.517 Percent Identity: 48.276

alignment_block:

US-09-049-696-16 x HCK_HUMAN

Align seg 1/1 to: HCK_HUMAN from: 1 to: 526

11 CCACGACTCCGCCAGACACCTAGTCCTGATCAACGCTCTGCTCTTG 60
 ||||| ||| ::|||::||| ::|||
 46 ProHisCysProValTyValProAspProThrSerThrIleLysProGI 62
 61 TCCTATATTCATCATCAACGACCATTCCTGGCATT 97
 ||||| ||| |||||::| |||||
 62 yProAsnSerHis...AsnSerAsnThrProGlyIle 73

seq_name: SwissProt_39:GUN6_DICDI

seq_documentation_block:

ID GUN6_DICDI STANDARD; PRT; 705 AA.
 AC P22699;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (SPORE
 DE GERMINATION PROTEIN 270-6) (CELLULOSE).
 GN CELA.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91002566; PubMed=1976383;
 RA Giordano R., Ohmachi T., Shaw D.R., Ennis H.L.;

RT "A shared internal threonine-glutamic acid-threonine-proline repeat
 RT defines a family of dictyostelium discoideum spore germination
 RT specific proteins.";
 RL Biochemistry 29:7264-7269(1990).
 RN [2]
 RP FUNCTION.
 RX MEDLINE=91332071; PubMed=1869562;
 RA Blume J.E., Ennis H.L.;
 RA "A dictyostelium discoideum cellulase is a member of a spore
 RT germination-specific gene family.";
 RT J. Biol. Chem. 266:15432-15437(1991).
 CC 1- FUNCTION: MAY DIGEST THE SPORE CELL WALL DURING GERMINATION, TO
 CC RELEASE THE ENCLOSED AMOEBAS.
 CC 1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC 1- DEVELOPMENTAL STAGE: FOUND PREDOMINANTLY DURING EARLY STAGES OF
 CC SPORE GERMINATION.
 CC 1- SIMILARITY: BELONGS TO CELLULOSE FAMILY E (FAMILY 9 OF GLYCOSYL
 CC HYDROLASES).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M33861; AAA52077.1; -
 CC PIR; A35621; A35621.
 CC HSP; P26221; 3TF4.
 CC Dictydb; DD04004; celA.
 DR InterPro; IPR001701; Glyco_hydro_9.
 DR Pfam; PF00759; Glyco_hydro_9; 1.
 DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
 DR PROSITE; PS00698; GLYCOSYL_HYDROL_F9_2; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein;
 KW Repeat; Sporulation; Signal.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 705 ENDOGLUCANASE.
 FT ACT_SITE 390 390 BY SIMILARITY.
 FT ACT_SITE 428 428 BY SIMILARITY.
 FT ACT_SITE 437 437 BY SIMILARITY.
 FT DOMAIN 463 552 THR-PRO REPEATS ("HINGE") (PRO-THR BOX).
 FT SIMILAR 463 552 TO THE T-E-T-P TANDEM REPEATS OF SPORE
 FT CARBOHYD 75 75 GERMINATION PROTEIN 270-II.
 FT SEQUENCE 705 AA; 75910 MW; F0457BDC07576A15 CRC64;

alignment_scores:
 Quality: 62.50 Length: 58
 Ratio: 1.953 Gaps: 3
 Percent Similarity: 55.172 Percent Identity: 32.759

alignment_block:

US-09-049-696-16 x GUN6_DICDI

Align seg 1/1 to: GUN6_DICDI from: 1 to: 705

8 CPTCCACACTCCGCCAGACACCT.....AGTCCTGATGAACGTC 51
 ||| ::||| |||||
 524 ProThrGluThrProThrGluThrProThrGluThrProThrGluThrVa 540
 52 TGTCTCTTCCTTAATTCATATCAACAGCACCATTCCTGGCATTCA 101
 ::||| |||::|
 540 lThrProThrProThrValThrProThrGluThrProSerSerGlyGlu 557
 102 TTTTAAATATG.....TGAAG 121
 |||::|
 557 erLeuSerIleTyLysSerGlyLeuLysAsnAspPheGlnAspTrpSer 573
 122 TGGATAGGAGAACTGCAGCTGTCA 145


```
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
KW Oxidoreductase; Peroxidase; Heme; Transmembrane; Glycoprotein;
KW Alternative splicing; Signal; Disease mutation; Polymorphism;
KW EGF-like domain.
FT SIGNAL 1 14
FT CHAIN 15 933
FT DOMAIN 796 839
FT ACT_SITE 239 239
FT ACT_SITE 396 396
FT ACT_SITE 494 494
FT TRANSMEM 847 871
FT CARBOHYD 129 129
FT CARBOHYD 307 307
FT CARBOHYD 342 342
FT CARBOHYD 569 569
FT DISULFID 800 814
FT DISULFID 808 823
FT DISULFID 825 838
FT VARSPIC 534 590
FT VARSPIC 874 933
FT VARIANT 257 257
FT VARIANT 373 373
FT VARIANT 398 398
FT VARIANT 453 453
FT VARIANT 725 725
FT VARIANT 799 799
FT CONFLICT 354 354
FT CONFLICT 371 371
FT CONFLICT 381 381
FT CONFLICT 574 574
FT CONFLICT 748 748
FT CONFLICT 847 847
FT CONFLICT 872 872
SQ SEQUENCE 933 AA; 102930 MW; 2363D2E5E80FA41 CRC64;

alignment_scores:
  Quality: 62.00 Length: 45
  Ratio: 2.583 Gaps: 2
Percent Similarity: 53.333 Percent Identity: 33.333

alignment_block:
US-09-049-696-16 x PERT_HUMAN ..
Align seg 1/1 to: PERT_HUMAN from: 1 to: 933
2 TTTATTCTCCACAGACTCCG.....CCAGAGACACCTAGTCC 39
|||||
366 PheValProPheArgAlaProAlaAlaCysAlaProGluProGlyIlePr 382
1
40 TGATGAACAGCTCTCTCTCTCT.....CCTAATATTATCATATCAACAGCA 83
|||||
382 oGlyGluThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerG 399
1
84 CCATCTCGGATTCACATTTTAAATAATTATGTGG 118
|||||
399 luValProSerLeuThrAlaLeuHisThrLeuTrp 410
|||||
seq_name: SwissProt_39.N551_SOYBN

seq_documentation_block:
ID N551_SOYBN STANDARD; PRT; 137 AA.
AC Q05544;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE EARLY NODULIN 55-1 PRECURSOR (N-55-1) (FRAGMENT).
GN EN055-1.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC euSids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. WILLIAMS; TISSUE=Root;
RX MEDLINE=94003074; PubMed=8400132;
RA deBlank C., Mylona P., Katinakis P.C., Bisseling T., Franssen H.;
RT Characterization of the soybean early nodulin cDNA clone Gmenod55.;
RL Plant Mol. Biol. 22:1167-1171(1993).
CC -!- SUBCELLULAR LOCATION: PERIBACTEROID MEMBRANE (POTENTIAL).
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AT EARLY STAGES OF NODULE
CC DEVELOPMENT. MAXIMAL EXPRESSION IS SEEN IN NODULES FROM 14-DAY-OLD
CC PLANTS AFTER WHICH LEVELS DECREASE.
CC -!- INDUCTION: DURING NODULATION IN LEGUME ROOTS AFTER RHIZOBIUM
CC INFECTION, AND AFTER RELEASE OF BACTERIA FROM THE INFECTION
CC THREAD.
CC -!- SIMILARITY: CONTAINS 1 PLASTOCYANIN-LIKE DOMAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X69156; CAA48908.1; ..
CC InterPro; IPR003245; Cu_bind_like.
CC Pfam; PF02298; Cu_bind_like; 1.
CC ProDom; PD003122; Cu_bind_like; 1.
KW Nodulation; Glycoprotein; Nitrogen fixation; Signal; Membrane.
FT NON_TER 1 1
FT SIGNAL <1 2
FT CHAIN ? 137
FT DOMAIN ? ?
FT CARBOHYD 13 13
FT CARBOHYD 51 51
FT CARBOHYD 68 68
FT SEQUENCE 137 AA; 14857 MW; 988DE6F89A2E0BE0 CRC64;

alignment_scores:
  Quality: 61.00 Length: 57
  Ratio: 2.033 Gaps: 2
Percent Similarity: 52.632 Percent Identity: 24.561

alignment_block:
US-09-049-696-16 x N551_SOYBN ..
Align seg 1/1 to: N551_SOYBN from: 1 to: 137
11 CCACAGACTCCGCCAGACACCTAGTCTGTATGAACCTGTCTCTCTG 60
|||||
80 ProSerProSerProSerProSerProSerProSerProSerProSe 96
|||||
61 TCCATAATTATCATATCAACACACCATTCCT..... 91
|||||
96 rProSerProSerLeuSerProSerProSerProSerProLeuProAsnAsnG 113
|||||
92 .....GGCATTCACATTTTAAATAATTATGTGGAAG 121
|||||
113 lnGlyValThrArgSerSerGlyAlaGluPheIleGlyValMet...Met 128
```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; S56585; AAB19916.1; -
DR EMBL; X53779; CAA37795.1; -
DR EMBL; M37890; AAA37234.1; -
DR EMBL; X59592; CAA42160.1; -
DR PIR; A35895; A35895.
DR PIR; A37255; A37255.
DR PIR; A37908; A37908.
DR PIR; S17198; S17198.
DR PIR; S34398; S34398.
DR HSSP; P06536; IRGD.
DR TRANSFAC; T00041; -
DR MGD; MGI:88064; Ar.
DR InterPro; IPR001103; Androgen_recep.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001628; zf-C4.
DR Pfam; PF02166; Androgen_recep; 1.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR PRINTS; PR00521; ANDROGENR.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00399; ZnF_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Steroid-binding.
FT DOMAIN 1 537 MODULATING (BY SIMILARITY).
FT NUCLEAR_BIND 539 604 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 539 559 C4-TYPE.
FT ZN_FING 575 599 C4-TYPE.
FT DOMAIN 670 899 LIGAND-BINDING.
FT DOMAIN 63 67 POLY-ARG.
FT DOMAIN 174 193 POLY-GLN.
FT DOMAIN 367 373 POLY-PRO.
FT DOMAIN 391 397 POLY-ALA.
FT DOMAIN 441 447 POLY-GLY.
SQ SEQUENCE 899 AA; 98193 MW; FD9EE07C07F7A568 CRC64;

alignment_scores:
Quality: 61.00 Length: 29
Ratio: 3.211 Gaps: 0
Percent Similarity: 65.517 Percent Identity: 37.931

alignment_block:

US-09-049-696-16 x ANDR_MOUSE ..

Align seg 1/1 to: ANDR_MOUSE from: 1 to: 899

11 CCACAGACTCCGCCAGACACCTAGTCTGATGAACGTCTGCTCTTG 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
125 ProGlnGlnProAlaProAlaProAspGlnAspSerAlaAlaProSe 141
61 TCCTAATATTTCATATCACACACCATTCCTGGCATTT 97
141 rThrLeuSerLeuLeuGlyProThrPheProGlyLeu 153

This Page Blank (uspto)

About: Results were produced by the GenCore software, version 4.5, Copyright (c) 1993-2000 CompuGen Ltd.

```

MODEL=frame-n2p_model DEV=xlp
Q/cgn2_1/USFT0_pool/US0304369
DB=Pir_68 -OFMT=fastan -SUFFIX=
-MINMATCH=0.150 -LOOPL=0.000 -L
-QGAPEX=0.050 -XGAPO=10.000 -L
-FGAPEX=7.000 -YGAPO=10.000 -Y
-DELETE_7.000 -START_1 -MATRIX=b
LIST=45 -DCALIGN=200 -TR_SCORE=
ALIGN=15 -MODE=LOCAL -OUTFMT=pfi
MAXLEN=200000000 -USER=US090496
LONGLOG NO XLPY -WAIT

```

Search time (sec): 470.790000

[illegible]

seq_documentation_block:
gob-5 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
C:Accession: JG0168
R:Komiyā, T.; Tanigawa, Y.; Hirohashi, S.
Biochem. Biophys. Res. Commun. 255, 347-351, 1999
A:Title: Cloning and Identification of the gene gob-5, which is expressed in intestinal
A:Reference number: JG0168; MUID:99160866
A:Accession: JG0168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-913 <KOM>
A:Cross-references: DDBJ:AB016592

alignment_scores:		
Quality:	179.50	Length: 49
Ratio:	4.378	Gaps: 1
Percent Similarity:	83.673	Percent Identity: 61.224

US-09-049-696-16 x JG0168

Align seq 1/1 CO: CCG0168 Locus: 1 LO: 913

2 TTTTATCTCCACAGACTCGGCAGAGACCACTAGTCCTGATGAACGTC 51
||||| |||| |
865 PheilerAalacInclupro.....ProileProgluaoserth 878

52 TGCTCCCTGGTGCCTAAATATTCATATCAACAGCAGCACCAATTCTGGCATTACA 101
: |||||:::|||||
878 rProCysProAspIleSerIleAsnSerThrIleProGlyIleHisV 895
: |||||:::|||||
102 TTTTAAAAATTATGTGGAGTGGATAGGAGAACTGCAGCTGCAATA 148
: |||||:::|||||
895 allLeuIysIleMetTrpLysTrpLeuClgVLuMetGlnValThrLeu 910
: |||||:::|||||

seq_name: PT12: A34035

DNA repair protein Xrcc-1 - mouse

C.Species: Mus musculus (house mouse)
C.Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 05-Nov-1999
C.Accession: A54659
R.Brookman, K.W.: Tebbis, R.S.; Allen, S.A.; Tucker, J.D.; Swiger, R.R.; Lam
A.Title: Isolation and characterization of mouse Xrcc-1, a DNA repair gene
A.Reference number: A54659; MUID:95048367

A: Molecule type: mRNA

A: Cross-references: GB:U

A;Gene: Xrcc-1

THE UNIVERSITY OF CHICAGO

Quality:	69.00	Length:	28
Ratio:	3.450	Gaps:	0
Percent Similarity:	71.429	Percent Identity:	46.429

US-09-049-696-16 x A54659

C: Genetics:

C:Superfamily: wheat insertion sequence WIS1 hypothetical 16K protein

```
alignment_scores:
  Quality: 64.50      Length: 36
  Ratio: 3.225       Gaps: 1
Percent Similarity: 55.556      Percent Identity: 38.889

alignment_block:
US-09-049-696-16 x S10084 ..
  Align seg 1/1 to: S10084 from: 1 to: 151
      8 CCTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAACGTCCTCC 57
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      63 ProProProProProProProProProProSerProSerIleSerLeuLeuPr 79
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      58 TTCT.....CCTAATATTTCATATCAACGACACCATTCCTG 92
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      79 opheProPheSerAspGlyProArgLeuAlaLeuAsnProValIleProA 96
      93 GCATTCAC 100
      |||
      96 spAspHis 98
seq_name: pir2.i49140
```

```
seq_documentation_block:
p62 ras-GAP associated phosphoprotein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C:Accession: I49140
C:Richard, S.; Yu, D.; Blumer, K.J.; Hausladen, D.; Olszowy, M.W.; Connelly, P.A.; Shaw,
Mol. Cell. Biol. 15, 186-197, 1995
A:Title: Association of p62, a multifunctional SH2- and SH3-domain-binding protein, with
A:Reference number: I49140; MUID:95097990
A:Accession: I49140
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-443 <RES>
A:Cross-references: EMBL:U17046; NID:G608527; PIDN:AAA64997.1; PID:G608528
C:Keywords: phosphoprotein
```

```
alignment_scores:
  Quality: 64.50      Length: 29
  Ratio: 3.395       Gaps: 1
Percent Similarity: 65.517      Percent Identity: 48.276

alignment_block:
US-09-049-696-16 x I49140 ..
  Align seg 1/1 to: I49140 from: 1 to: 443
      8 CPTCCACAGACTCCGCCAGAGACACCTAGTCTGATGAACG..... 49
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      64 ProProLeuLeuProProSerThrProGlyProAspAlaThrValValGI 80
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      50 .CTGTCTCTTGTCTTAATTTCATATCAACGACAC 85
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      80 ySerAlaProThrProLeuLeuProProSerAlaThr 92
seq_name: pir2.A40357
```

```
seq_documentation_block:
homeotic protein lab1 - African clawed frog (fragment)
C:Species: Xenopus laevis (African clawed frog)
C:Date: 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change 24-Sep-1999
C:Accession: A40357; C40856
R:Siye, H.L.; Cheng, P.F.
Genes Dev. 5, 1321-1332, 1991
A:Title: Retinoic acid perturbs the expression of Xhox lab genes and alters mesodermal d
A:Reference number: A40357; MUID:91331304
```

```
A:Accession: A40357
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-272 <SIV>
A:Cross-references: GB:X62052; NID:g396598; PIDN:CAA43980.1; PID:g396599
R:Blumberg, B.; Wright, C.V.E.; De Robertis, E.M.; Cho, K.W.Y.
Science 253, 194-196, 1991
A:Title: Organizer-specific homeobox genes in Xenopus laevis embryos.
A:Reference number: A40856; MUID:91305940
A:Accession: C40856
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 126-141 'Q', 143-185 <BLU>
A:Cross-references: GB:M63873
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:127-183/Domain: homeobox homology <BOX>
```

```
alignment_scores:
  Quality: 64.00      Length: 63
  Ratio: 1.882       Gaps: 3
Percent Similarity: 53.968      Percent Identity: 34.921

alignment_block:
US-09-049-696-16 x A40357 ..
  Align seg 1/1 to: A40357 from: 1 to: 272
      14 CAGACTCGCCAGAGACACCTAGTCTGTGATGAA.....ACGTCTCC 54
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      109 ArgAsnProProLysLysSerIleProSerGlyThrGlyValThrSerPr 125
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      55 TCCTGTCTCTATATCATATCATATCAACGACCATTCCTGCGATTCA.... 99
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      125 opProcys...AsnValArgThrAsnPheThrThrLysGlnLeuThrGluL 141
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      100 .....CATTTTAAATAATTATGTGGAAGTGTGATAGGAGAACT 135
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      141 euGlyLysGluPheHisPheAsnLysTyrLeuThrArgAlaArgIle 157
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      135 GCAGCTGCTCAATAGCTAGGCTGGAATTTTGTCTCAGATA 174
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      158 GluIleAlaAsnSerLeuGlnLeuAsnAspThrGlnVal 170
```

```
seq_name: pir2.S21961

seq_documentation_block:
proline-rich protein APG - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997
C:Accession: S21961
R:Roberts, M.R.; Foster, G.D.; Blundell, R.P.; Robinson, S.W.; Draper, J.; Scott, R.J
submitted to the EMBL Data Library, August 1991
A:Description: Cloning and characterization of a proline-rich gene expressed specific
A:Reference number: S16748
A:Accession: S21961
A:Molecule type: DNA
A:Residues: 1-534 <ROB>
A:Cross-references: EMBL:X60377; NID:g22598; PID:g22599
C:Genetics:
A:Gene: APG
A:Introns: 256/1; 299/3; 387/3; 470/1
```

```
alignment_scores:
  Quality: 64.00      Length: 37
  Ratio: 3.048       Gaps: 1
Percent Similarity: 56.757      Percent Identity: 37.838

alignment_block:
US-09-049-696-16 x S21961 ..
```

Align seg 1/1 to: S21961 from: 1 to: 534

5 ATTCTCCACAGACTCCGCCA.....GAGACACCTAG 36
:::|||||::: |||||
169 ValProProHisGlyProProLysProAlaProAlaProThrProAl 185
:::|||||::: |||||
37 TCCTGATGAACGCTGCTCTGTCCTTAATATCATATCAACAGCACCA 86
:::|||||::: |||||
185 aProSerProLysProAlaProSerProProLysProGluAsnLysThrI 202
:::|||||::: |||||

87 TTCCTGGCATT 97
|||||:::|||||

202 leProAlaVal 205
:::|||||::: |||||

seq_name: pir2:T05264

seq_documentation_block:

probable serine/threonine-specific protein kinase (EC 2.7.1.-) T4L20.20 - Arabidopsis th

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 17-Mar-2000

C:Accession: T05264

R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De
ewes, H.W.; Mayer, K.F.X.; Schueller, C.

submitted to the Protein Sequence Database, September 1998

A:Reference number: Z15406

A:Accession: T05264

A:Molecule type: DNA

A:Residues: 1-674 <BEV>

A:Cross-references: EMBL:AL023094

A:Experimental source: cultivar Columbia; BAC clone T4L20

C:Genetics:

A:Map position: 4

A:Introns: 180/1; 221/1; 381/1; 421/1; 444/3; 470/2; 518/3; 583/3

A:Note: T4L20.20

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

C:keywords: phosphotransferase; protein kinase

alignment_scores:

Quality:	64.00	Length:	27
Ratio:	3.368	Gaps:	0
Percent Similarity:	70.370	Percent Identity:	44.444

alignment_block:

US-09-049-696-16 x T05264 ..

Align seg 1/1 to: T05264 from: 1 to: 674

11 CCACAGACTCCGCCAGACACACTAGTCTGATGAACGCTGCTCTTG 60
|||||:::|||||::: |||||

32 ProProThrProProSerProProSerSerIleSerAlaProPr 48
:::|||||::: |||||

61 TCCTAATATTCATATCAACAGCACCATTCT 91
:::|||||::: |||||

48 oProAspIleSerAlaSerPheSerProPro 58
:::|||||::: |||||

seq_name: pir2:F86387

seq_documentation_block:

probable pto kinase interactor [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: F86387

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: F86387

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-760 <STO>

A:Cross-references: GB:AE005172; NID:g11079512; PIDN:AAG29223.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

alignment_scores:

Quality:	64.00	Length:	19
Ratio:	4.571	Gaps:	0
Percent Similarity:	73.684	Percent Identity:	52.632

alignment_block:

US-09-049-696-16 x F86387 ..

Align seg 1/1 to: F86387 from: 1 to: 760

18 CCTCCACAGACTCCGCCAGACACACTAGTCTGATGAACGCTGCTCC 57
|||||:::|||||::: |||||

150 ProProGluSerProProSerLeuProAlaProAspProProSerAsnPr 166
:::|||||::: |||||

50 TTGTCTCT 64
:::|||||::: |||||

165 oLeuPro 168
:::|||||::: |||||

seq_name: pir2:A86335

seq_documentation_block:

hypothetical protein AAF79900.1 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: A86335

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: A86335

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1137 <STO>

A:Cross-references: GB:AE005172; NID:g8778985; PIDN:AAF79900.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

alignment_scores:

Quality:	64.00	Length:	37
Ratio:	3.048	Gaps:	1
Percent Similarity:	56.757	Percent Identity:	37.838

alignment_block:

US-09-049-696-16 x A86335 ..

Align seg 1/1 to: A86335 from: 1 to: 1137

5 ATTCTCCACAGACTCCGCCA.....GAGACACCTAG 36
:::|||||::: |||||

170 ValProProHisGlyProProLysProAlaProAlaProThrProAl 195
:::|||||::: |||||

37 TCCTGATGAACGCTGCTCTGTCCTTAATATTCATATCAACAGCACCA 86
:::|||||::: |||||

195 aProSerProLysProAlaProSerProProLysProGluAsnLysThrI 212
:::|||||::: |||||

87 TTCTTGGCATT 97
|||||:
212 leProAlaVal 215

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,269C
FILING DATE: 02-Jun-95
ATTORNEY/AGENT INFORMATION:
NAME: Lebovitz, Richard M.
REGISTRATION NUMBER: 37,067
REFERENCE/DOCKET NUMBER: Popov-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 243-6333
TELEFAX: (703) 243-6410
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 910 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-460-269C-2

```
alignment_block:
US-09-049-696-16 x US-08-460-269C-2
..
```

3 CCTCCAGAGACTCGCCAGAGACCTAGTCTGAT..... 43

604 gGluLeuSerAla...AlaAlaAsnAlaAlaValAsnThr.....GlyG 618
03 CCGTTCACAGCTGGTTGTATTAATAAATTAACTCCTCC 118

119 AAGTGGATAGGAGAAGCTGCAGCTGTCA 145
:::
635 LysArgLeuGlyGluLeuArgLeuAsn 643

```
seq_documentation_block: US/08460269C
: Sequence 4, Application
```

APPLICANT: CLARE, JEFFREY J.
ROMANOS, MICHAEL A.
WIFE OF TITANUS

ADDRESSEE: Millen, White, Zelano & Branigan, P.C.

STATE: VA
COUNTRY: USA
ZIP: 22201

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```
; .  
; .  
; .  
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
alignment_block:
```

2 TTTATTCTCCACAGACTCGCCAGACACACCTAGTCTGATGAACGTC 51
 |||
 209 pheileProGlnThrProProGluThrProSerProAspGluThrSe 225

225 rAlaProCys 228

```
seq_documentation_block:
; Sequence 2, Application US/08460269C
```

APPLICANT:	CLARE, JEFFREY J.
ROMANOS, MICHAEL A.	
TITLE OF INVENTION:	EXPRESSION OF HETEROLOGOUS PROTEINS

; ;
CORRESPONDENCE ADDRESS:
ADDRESSEE: Millen, White, Zelano & Branigan, P
00

COUNTRY: USA
ZIP: 22201

; COMPUTER READABLE FORM:


```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,269C
; FILING DATE: 02-Jun-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: Popov-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333
; TELEFAX: (703) 243-6410
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 911 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-08-460-269C-4
alignment_scores:
  Quality: 63.00      Length: 59
  Ratio: 1.969        Gaps: 4
  Percent Similarity: 54.237      Percent Identity: 32.203
alignment_block:
  US-09-049-696-16 x US-08-460-269C-4 ..
  Align seg 1/1 to: US-08-460-269C-4 from: 1 to: 911
      8 CCTCCAGACTCCGCGAGACACCTAGTCCTGAT..... 43
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
589 ProProGlnArgGlnProGluAlaProAlaProGlnProProAlaGlyAr 605
      44 .GAACGCTGCTCTCTCTCTCTATATTCATATCAACAGCACCATTCTCG 92
      ||| ||||| ||| ||| ||| ||||| ||||| |||||
605 gGluLeuSerAla...AlaAlaAsnAlaAlaValAsnThr.....GlyG 619
      93 GCATTACATTTTAAAAATTATGTGG..... 118
      ||||| ||| ||| ||| ||| ||||| ||||| |||||
619 lyValGlyLeuAlaSerThrLeuTrpTyrAlaGluSerAsnAlaLeuSer 635
      119 AAGTGGATAGGAGAACTGCAGCTGTCA 145
      ||| ||||| ||||| ||||| ||||| ||||| |||||
636 LysArgLeuGlyGluLeuArgLeuAsn 644
seq_name: /cgn2_6/ptodata/2/1aa/68_COMB.pep:US-08-460-269C-6
seq_documentation_block:
; Sequence 6, Application US/08460269C
; Patent No. 6157548
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; ROMANOS, MICHAEL A.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,269C
; FILING DATE: 02-Jun-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lebovitz, Richard M.
; REGISTRATION NUMBER: 37,067
; REFERENCE/DOCKET NUMBER: Popov-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 243-6333
; TELEFAX: (703) 243-6410
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 922 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-08-460-269C-6
alignment_scores:
  Quality: 63.00      Length: 59
  Ratio: 1.969        Gaps: 4
  Percent Similarity: 54.237      Percent Identity: 32.203
alignment_block:
  US-09-049-696-16 x US-08-460-269C-6 ..
  Align seg 1/1 to: US-08-460-269C-6 from: 1 to: 922
      8 CCTCCAGACTCCGCGAGACACCTAGTCCTGAT..... 43
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
600 ProProGlnArgGlnProGluAlaProAlaProGlnProProAlaGlyAr 616
      44 .GAACGCTGCTCTCTCTCTATATTCATATCAACAGCACCATTCTCG 92
      ||| ||||| ||| ||| ||| ||||| ||||| |||||
616 gGluLeuSerAla...AlaAlaAsnAlaAlaValAsnThr.....GlyG 630
      93 GCATTACATTTTAAAAATTATGTGG..... 118
      ||||| ||| ||| ||| ||| ||||| ||||| |||||
630 lyValGlyLeuAlaSerThrLeuTrpTyrAlaGluSerAsnAlaLeuSer 646
      119 AAGTGGATAGGAGAACTGCAGCTGTCA 145
      ||| ||||| ||||| ||||| ||||| ||||| |||||
647 LysArgLeuGlyGluLeuArgLeuAsn 655
seq_name: /cgn2_6/ptodata/2/1aa/PCTUS95-05008-17
seq_documentation_block:
; Sequence 17, Application PC/TUS9505008
; GENERAL INFORMATION:
; APPLICANT: Sugen, Inc.
; APPLICANT: 515 Galveston Drive
; APPLICANT: Redwood City, California 94063-4720
; APPLICANT: United States of America
; APPLICANT: Wissenschaften E.V.
; APPLICANT: Hofgarten Str. 2
; APPLICANT: Munchen 80539
; APPLICANT: Germany
; TITLE OF INVENTION: Novel Megakaryocytic Protein Tyrosine
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05008
; FILING DATE: 24-APR-1995
```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/232,545
; FILING DATE: 22-APR-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7683-074
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212)869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; PCT-US95-05008-17

alignment_scores:
  Quality: 62.50      Length: 29
  Ratio: 3.289       Gaps: 1
  Percent Similarity: 65.517      Percent Identity: 48.276

alignment_block:
US-09-049-696-16 x PCT-US95-05008-17
..
Align seg 1/1 to: PCT-US95-05008-17 from: 1 to: 505

11 CCACGACTCCGCGCAGACACCTAGTCTGATGAACGTCGTCTCTTG 60
||||| ||| :|||:|||||:||||| :||| |||
25 PROHiscysProValrYrValProAspProThrSerThrIleysProGl 41
||||| ||| |||||:| |||||
61 TCCTAATATTTCATATCAACAGCACCATTCTGGCATT 97
||||| ||| |||||:| |||||
41 yProAsnSerHis...AsnSerAsnThrProGlyIle 52
||||| ||| |||||:| |||||

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-217-327-4

seq_documentation_block:
; Sequence 4, Application US/08217327
; Patent No. 5474925
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E
; APPLICANT: Barton, Kenneth A
; TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles and Brady
; STREET: P.O. Box 2113
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/217,327
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/812,233
; FILING DATE: 19-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386

```

```

; REFERENCE/DOCKET NUMBER: 1122990831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 214 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-217-327-4

alignment_scores:
  Quality: 61.00      Length: 30
  Ratio: 3.588       Gaps: 0
  Percent Similarity: 56.667      Percent Identity: 40.000

alignment_block:
US-09-049-696-16 x US-08-217-327-4
..
Align seg 1/1 to: US-08-217-327-4 from: 1 to: 214

3 CCTCCAGACTCCGCGCAGACACCTAGTCTGATGAACGTCGTCTGCC 57
||||| |||||:|||||:||||| :||| |||
124, ProProAlaThrProProAlaThrProProAlaThrProProAlaThrProPro 140
||||| |||||:|||||:||||| :||| |||
58 TTGTCCTAATATTTCATATCAACAGCACCATTCTGGCATT 97
||||| ||| |||||:| |||||
140 oAlaProLeuAlaSerProProAlaThrValProAlaIle 153
||||| ||| |||||:| |||||

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-313-200-1

seq_documentation_block:
; Sequence 1, Application US/08313200
; Patent No. 5998153
; GENERAL INFORMATION:
; APPLICANT: Baker, James R.
; APPLICANT: Koenig, Ronald J.
; TITLE OF INVENTION: THYROID PEROXIDASE EPITOPIC REGIONS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,200
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kouski, Antoinette F.
; REGISTRATION NUMBER: 34,202
; REFERENCE/DOCKET NUMBER: 20344-20658.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 933 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO

```

```

OTHER INFORMATION: binding fusion protein"
FEATURE:
  NAME/KEY: Region
  LOCATION: 457..517
  OTHER INFORMATION: /note= "non-reactive fragment"
FEATURE:
  NAME/KEY: Region
  LOCATION: 457..633
  OTHER INFORMATION: /note= "TPO region within fusion plasmid pMalTPO"
  OTHER INFORMATION:
FEATURE:
  NAME/KEY: Region
  LOCATION: 457..933
  OTHER INFORMATION: /note= "TPO binding region within plasmid pMalTPO"
  OTHER INFORMATION:
FEATURE:
  NAME/KEY: Region
  LOCATION: 465..933
  OTHER INFORMATION: /note= "TPO binding region of maltose binding region fusion construct"
  OTHER INFORMATION:
FEATURE:
  NAME/KEY: Region
  LOCATION: 513..633
  OTHER INFORMATION: /note= "recombinant TPO"
FEATURE:
  NAME/KEY: Region
  LOCATION: 517..630
  OTHER INFORMATION: /note= "TPO binding or epitopic region"
  OTHER INFORMATION:
FEATURE:
  NAME/KEY: Region
  LOCATION: 517..633
  OTHER INFORMATION: /note= "TPO binding or epitopic region"
FEATURE:
  NAME/KEY: Region
  LOCATION: 573..633
  OTHER INFORMATION: /note= "TPO binding or epitopic region"
  OTHER INFORMATION:
FEATURE:
  NAME/KEY: Region
  LOCATION: 590..611
  OTHER INFORMATION: /note= "TPO region within maltose binding fusion protein"
FEATURE:
  NAME/KEY: Region
  LOCATION: 590..615
  OTHER INFORMATION: /note= "TPO binding or epitopic region"
FEATURE:
  NAME/KEY: Binding-site
  LOCATION: 590..675
  OTHER INFORMATION: /note= "TPO binding or epitopic region"
  OTHER INFORMATION:
FEATURE:
  NAME/KEY: Region
  LOCATION: 592..613
  OTHER INFORMATION: /note= "TPO binding or epitopic region"
  OTHER INFORMATION:
FEATURE:
  NAME/KEY: Region
  LOCATION: 596..611
  OTHER INFORMATION: /note= "Tpo region within fusion protein"
  OTHER INFORMATION:
FEATURE:
  NAME/KEY: Region
  LOCATION: 602..615
  OTHER INFORMATION: /note= "TPO region containing divergent sequences"
  OTHER INFORMATION:
FEATURE:
  NAME/KEY: Region
  LOCATION: 611..615
  OTHER INFORMATION: /note= "TPO binding or epitopic region"
  OTHER INFORMATION:

```

OTHER INFORMATION: region"
FEATURE: Region
NAME/KEY: Region
LOCATION: 631..933
OTHER INFORMATION: /note= "TPO binding or epitopic"
OTHER INFORMATION: region"
FEATURE: Region
NAME/KEY: Region
LOCATION: 632..933
OTHER INFORMATION: /note= "TPO region within maltose
OTHER INFORMATION: binding fusion protein"
FEATURE: Region
NAME/KEY: Region
LOCATION: 633..768
OTHER INFORMATION: /note= "TPO binding or epitopic"
alignment_scores:
Quality: 61.00 Length: 45
Ratio: 2.652 Gaps: 2
Percent Similarity: 51.111 Percent Identity: 33.333
alignment_block:
US-09-049-696-16 x US-08-313-200-1
Align seg 1/1 to: US-08-313-200-1 from: 1 to: 933
2 TTTATTCCTCCACAGACTCCG.....CCAGAGACCTAGTCC 39
|||||:|||||: |||
366 PheValProArgProAlaAlaCysAlaProGluProGlyIlePr 382
40 TGATGAACGTCCTCTGT.....CCTATATTCATATCAACAGCA 83
|||||: |||
382 OGlyGluThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerG 399
84 CCATTCTGGCATTCACATTTTAAATAATATGTGG 118
399 luValProSerLeuThrAlaLeuHisThrLeuTrp 410
seq_name: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US93-03837-1
seq_documentation_block:
Sequence 1, Application PC/TUS9303837
GENERAL INFORMATION:
APPLICANT: Baker, Jr., James R.
APPLICANT: Koenig, Ronald J.
TITLE OF INVENTION: Thyroid Peroxidase Epitopic Regions
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: USA
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/03837
FILING DATE: 19930422
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Lewak, Anna M.
REGISTRATION NUMBER: 33006
REFERENCE/DOCKET NUMBER: 2115-00658PPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (313) 641-1600
TELEFAX: (313) 641-0270
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 933 amino acids

TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
DEVELOPMENTAL STAGE: Mature
TISSUE TYPE: Thyroid gland (from people with Grave's
IMMEDIATE SOURCE:
CLONE: pHTPO-2.8
PUBLICATION INFORMATION:
AUTHORS: Kimura, S.
AUTHORS: Kotani, T.
AUTHORS: McBride, O. W.
AUTHORS: Umeki, K.
AUTHORS: Nakayama, T.
AUTHORS: Ohtaki, S.
AUTHORS: Hirai, K.
TITLE: Human thyroid peroxidase: Complete cDNA and
TITLE: protein sequence, chromosome mapping, and
TITLE: identification of two alternately spliced mRNAs
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 84
PAGES: 5555-5559
DATE: 1987
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 3048
PCT-US93-03837-1
alignment_scores:
Quality: 61.00 Length: 45
Ratio: 2.652 Gaps: 2
Percent Similarity: 51.111 Percent Identity: 33.333
alignment_block:
US-09-049-696-16 x PCT-US93-03837-1
Align seg 1/1 to: PCT-US93-03837-1 from: 1 to: 933
2 TTTATTCCTCCACAGACTCCG.....CCAGAGACCTAGTCC 39
|||||:|||||: |||
366 PheValProArgProAlaAlaCysAlaProGluProGlyIlePr 382
40 TGATGAACGTCCTCTGT.....CCTATATTCATATCAACAGCA 83
|||||: |||
382 OGlyGluThrArgGlyProCysPheLeuAlaGlyAspGlyArgAlaSerG 399
84 CCATTCTGGCATTCACATTTTAAATAATATGTGG 118
399 luValProSerLeuThrAlaLeuHisThrLeuTrp 410
seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-922-635-22
seq_documentation_block:
Sequence 22, Application US/08922635A
Patent No. 6033871
GENERAL INFORMATION:
APPLICANT: PILETZ, John E.
APPLICANT: IVANOV, Tina R.
TITLE OF INVENTION: DNA MOLECULES ENCODING IMIDALINE RECEPTIVE POLYPEPTIDES
TITLE OF INVENTION: AND POLYPEPTIDES ENCODED THEREBY
FILE REFERENCE: Corrected Sequence Listing
Patent No. 6033871
CURRENT APPLICATION NUMBER: US/08/922,635A
CURRENT FILING DATE: 1997-09-03
EARLIER APPLICATION NUMBER: 08/650,766
EARLIER FILING DATE: 1996-05-20
EARLIER APPLICATION NUMBER: 60/012,600
EARLIER FILING DATE: 1996-03-01
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.0

```
; SEQ ID NO 22
; LENGTH: 1070
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-922-635-22

alignment_scores:
  Quality: 61.00      Length: 69
  Ratio: 1.794        Gaps: 1
  Percent Similarity: 49.275  Percent Identity: 20.290

alignment_block:
US-09-049-696-16 x US-08-922-635-22 ..
Align seg 1/1 to: US-08-922-635-22 from: 1 to: 1070
5 ATTCTCCACAGACTCCGCCAGACACCTAGTCTGATGAAGCTGCG 54
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
652 ValProGluGluThrProValGluAlaProAlaProProAlaGluAl 668
55 TCCT..... 58
||||
668 aproAlaGlnTyrProSerGluHisLeuIleGlnAlaThrSerGluGluA 685
59 .....TGCTCTAATATTCATATCAACAGC 82
685 snGlnIleProSerHisLeuProAlaCysProSerLeuArgHisValAla 701
83 ACCATCTCGCATTCACATTTTAAATATTTGGAAAGTGATAGGAGA 132
:|||||:|||||:|||||:|||||:|||||:|||||:
702 SerLeuArgGlySerAlaIleGluLeuPheHisSerIleAlaGlu 718
133 ACTGCAG 139
:|||||:
718 uValGlu 720

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-141-212-8

seq_documentation_block:
; Sequence 8, Application US/09141212
; Patent No. 6200777
; GENERAL INFORMATION:
; APPLICANT: MACPHEE, COLIN
; APPLICANT: PATEL, LISA
; TITLE OF INVENTION: NOVEL COMPOUNDS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/141,212
; FILING DATE: 27-AUG-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 97306807.5
; FILING DATE: 01-SEP-1997
; APPLICATION NUMBER: EP 98300687.5
; FILING DATE: 30-JAN-1998
; APPLICATION NUMBER: GB 9807720.9
; FILING DATE: 08-APR-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031

REFERENCE/DOCKET NUMBER: GP-30012
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0700
TELEX: 846169
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
  LENGTH: 878 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: protein
US-09-141-212-8

alignment_scores:
  Quality: 60.50      Length: 31
  Ratio: 3.025        Gaps: 1
  Percent Similarity: 64.516  Percent Identity: 41.935

alignment_block:
US-09-049-696-16 x US-09-141-212-8 ..
Align seg 1/1 to: US-09-141-212-8 from: 1 to: 878
8 CCTCCACAGACTCCGCCAGACACCT.....AGTCTGTGATGAAC 48
:|||||:|||||:|||||:|||||:|||||:|||||:
848 ProProGlnThrProProAspLeuProAlaGlnAlaProAspLeuCy 864
49 GTCGTCTCTTGTCTCTAATATTCATATCAACAGCACCATTCTCT 91
:|||||:|||||:|||||:|||||:|||||:|||||:
864 sSerLeuLeuCysLeuProIleThrPheSerGlyAlaLeuPro 878

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-561-138-8

seq_documentation_block:
; Sequence 8, Application US/09561138
; Patent No. 6258580
; GENERAL INFORMATION:
; APPLICANT: MACPHEE, COLIN
; APPLICANT: PATEL, LISA
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30012-D1
; CURRENT APPLICATION NUMBER: US/09/561,138
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 09/141,212
; PRIOR FILING DATE: 1998-08-27
; PRIOR APPLICATION NUMBER: EP 97306807.5
; PRIOR FILING DATE: 1997-09-01
; PRIOR APPLICATION NUMBER: EP 98300687.5
; PRIOR FILING DATE: 1998-01-30
; PRIOR APPLICATION NUMBER: GB 9807720.9
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 878
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-561-138-8

alignment_scores:
  Quality: 60.50      Length: 31
  Ratio: 3.025        Gaps: 1
  Percent Similarity: 64.516  Percent Identity: 41.935

alignment_block:
US-09-049-696-16 x US-09-561-138-8 ..
Align seg 1/1 to: US-09-561-138-8 from: 1 to: 878
8 CCTCCACAGACTCCGCCAGACACCT.....AGTCTGTGATGAAC 48
```

|||||
848 ProProGlnThrProProAspLeuProAlaGlnAlaAlaProAspLeuCy 864
|||||

49 GTCTGCTCCTTGCTCAATATCATATCAACAGCACCATTCCT 91
|||||

864 sSerLeuLeuCyLeuProIleThrPheSerGlyAlaLeuPro 878
|||||

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-742-440A-6

seq_documentation_block:

; Sequence 6, Application US/08742440A

; Patent No. 5892014

; GENERAL INFORMATION:

; APPLICANT: Coughlin, Shaun

; APPLICANT: Ishihara, Hiroaki

; APPLICANT: Connolly, Andrew

; TITLE OF INVENTION: Protease Activated Receptor

; TITLE OF INVENTION: 3 and Uses Thereof

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bozicevic & Reed, LLP

; STREET: 285 Hamilton Avenue, Suite 200

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94301

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/742.440A

; FILING DATE: 30-OCT-1996

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Sherwood, Pamela J

; REGISTRATION NUMBER: 36,677

; REFERENCE/DOCKET NUMBER: UCAL/060PAT

; TELEPHONE: 650-327-3400

; TELEFAX: 650 327-3231

; TELEX:

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 408 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; FEATURE:

US-08-742-440A-6

alignment_scores:

Quality: 60.00 Length: 41

Ratio: 2.308 Gaps: 1

Percent Similarity: 63.415 Percent Identity: 31.707

alignment_block:

US-09-049-696-16/rev x US-08-742-440A-6 ..

Align seg 1/1 to: US-08-742-440A-6 from: 1 to: 408

225 CATTTTGAATAATTTTATAATCAAAAAAGGATGATTTATTTAT 176
|||||

344 H1sAlaAsnTyrtYrYrAsnAsnThrAspGly.....LeutyPheIl 358
|||||

175 TTATCTGACAAAAATTCAGCCCTAGGCTATTGACAGCTGCAGTTCTCCTA 126
|||||

358 eTyLeuIleAlaLeuCyLeuGlySerLeuAsnSerCysLeuAspProp 375
|||||

125 TCCACTTCCACATAATTTTAA 103
|||||

375 heLeuTyPheLeuMetSerLys 382
|||||

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-029-213B-25

seq_documentation_block:

; Sequence 25, Application US/09029213B

; Patent No. 6180098

; GENERAL INFORMATION:

; APPLICANT: CHRISTIAN, Peter D.

; TITLE OF INVENTION: RECOMBINANT HELICOVERPA BACULOVIRUSES

; TITLE OF INVENTION: EXPRESSING HETEROLOGOUS DNA

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: McDermott, Will & Emery

; STREET: 600 13th Street, NW

; CITY: Washington

; STATE: District of Columbia

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/029,213B

; FILING DATE: 31-AUG-1998

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Joseph Hyosuk Kim

; REGISTRATION NUMBER: 41,425

; REFERENCE/DOCKET NUMBER: 50179-048

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-756-8000

; TELEFAX: 202-756-8087

; INFORMATION FOR SEQ ID NO: 25:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 323 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-029-213B-25

alignment_scores:

Quality: 59.00 Length: 29

Ratio: 3.471 Gaps: 1

Percent Similarity: 58.621 Percent Identity: 41.379

alignment_block:

US-09-049-696-16 x US-09-029-213B-25 ..

Align seg 1/1 to: US-09-029-213B-25 from: 1 to: 323

5 ATTCCTCCACAGACTCGCCAGACACCTAGTCTGTGATGAACAGTCTGC 54
|||||

44 MetProProIleProIleProIleProProPro.....Pr 56
|||||

55 TCCTTCTCTTAATATTCATATCAACAGCACCATTCCT 91
|||||

56 oProProGlnThrAsnLeuSerThrIlePro 68
|||||

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-251-645-6

seq_documentation_block:

; Sequence 6, Application US/09251645

; Patent No. 6281413

; GENERAL INFORMATION:

; APPLICANT: Kramer, Vance C.

; APPLICANT: Morgan, Michael K.

```

; APPLICANT: Anderson, Arne R.
; APPLICANT: Hart, Hope
; APPLICANT: Warren, Gregory W.
; APPLICANT: Dunn, Martha
; APPLICANT: Chen, Jeng S.
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: GCG1963/A
; CURRENT APPLICATION NUMBER: US/09/251,645
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1584
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-251-645-6

```

```

alignment_scores:
    Quality: 58.00      Length: 29
    Ratio: 2.636        Gaps: 0
    Percent Similarity: 75.862  Percent Identity: 44.828

```

```

alignment_block:
US-09-049-696-16/rev x US-09-251-645-6 ..

Align seg 1/1 to: US-09-251-645-6 from: 1 to: 1584

90 GGAATGGTGTCTGTGATATGAATATTAGCAGCAGCAGCGTTTCATC 41
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1367 GlyIleIleValGlyIleValSerLeuGlyAlaGlyAlaAlaIleSerAl 1383

40 AGGACTAGGTGTCTCTGCGGAGTCTGTGGAGGAATA 4
|||||:|||||:|||||:|||||:|||||:|||||:
1383 aGlyLeuIleAlaAlaGlyAlaGlyAlaLeuGlyAlaIle 1395

```

This Page Blank (uspto)

CC treatment of these conditions. The ICACC proteins can be used:
CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
CC reagents, and (b) as therapeutic (Ab); (ii) as fragments or mutants, as
CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
CC identify modulators and binding partners. ICACC polynucleotides can be
CC used to generate transgenic animals or recombinant cells, used to screen
CC for antagonists, also as a source of therapeutic antisense agents or
CC diagnostic probes (for quantifying mRNA expression, e.g. for
CC identification of modulators).
XX
SQ Sequence 914 AA;

alignment_scores:
Quality: 278.00 Length: 50
Ratio: 5.560 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-16 x AAB74824 ..
Align seg 1/1 to: AAB74824 from: 1 to: 914
2 TTTATTCCTCCACAGACTCGCCAGACACCTAGTCTGTATGAACGTC 51
|||||
865 PheileProGlnThrProProGluThrProSerProAspGluThrSe 881
52 TGCTCCTGTGCTTAATATTCATATCAACAGACACCATTCCTGGCATTACACA 101
|||||
881 rAlaProCysProAsnIleHisIleAsnSerThrIleProGlyIleHisI 898
102 TTTTAAATATTCGAGAGTGGATAGAGAGACTGCGACCTGTCATATAGCC 151
|||||
898 leLeuLysIleMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914

seq_name: /SDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW24514

seq_documentation_block:

ID AAM24514 standard; Protein; 914 AA.
XX
AC AAM24514;
XX
DT 12-OCT-2001 (first entry)
XX
DE C902P predicted amino acid sequence.
XX
KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW Immunogenic; gene therapy; vaccine; colonic cancer.
XX
OS Homo sapiens.
XX
PN WO200149716-A2.
XX
PD 12-JUL-2001.
XX
PF 29-DEC-2000; 2000WO-US35596.
XX
PR 30-DEC-1999; 99US-0476296.
PR 10-JAN-2000; 2000US-0480321.
PR 15-FEB-2000; 2000US-0504629.
PR 06-MAR-2000; 2000US-0519444.
PR 19-MAY-2000; 2000US-0575251.
PR 29-JUN-2000; 2000US-0609448.
PR 28-AUG-2000; 2000US-0649811.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
XX WPI; 2001-441847/47.
XX
PT Colon tumor associated proteins and nucleic acids useful for the

PT prevention, diagnosis and treatment of colonic cancer -
XX
PS Claim 2; Page 440-443; 472pp: English.
XX
CC The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be
CC used to treat disorders associated with decreased expression by
CC rectifying mutations or deletions in a patient's genome that affect the
CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host
CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
XX given in the exemplification of the present invention.
SQ Sequence 914 AA;

alignment_scores:

Quality: 278.00 Length: 50
Ratio: 5.560 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-16 x AAM24514 ..
Align seg 1/1 to: AAM24514 from: 1 to: 914
2 TTTATTCCTCCACAGACTCGCCAGACACCTAGTCTGTATGAACGTC 51
|||||
865 PheileProGlnThrProProGluThrProSerProAspGluThrSe 881
52 TGCTCCTGTGCTTAATATTCATATCAACAGACACCATTCCTGGCATTACACA 101
|||||
881 rAlaProCysProAsnIleHisIleAsnSerThrIleProGlyIleHisI 898
102 TTTTAAATATTCGAGAGTGGATAGAGAGACTGCGACCTGTCATATAGCC 151
|||||
898 leLeuLysIleMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914
seq_name: /SDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAB73716

seq_documentation_block:

ID AAB73716 standard; Protein; 914 AA.
XX
AC AAB73716;
XX
DT 11-SEP-2001 (first entry)
XX
DE Human CLCAL protein, SEQ ID NO:2.
XX
KW Human CLCAL; goblet cell; mouse Gob-5 orthologue; drug screening;
KW expression inhibition; antisense therapy; gene therapy;
KW chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic.
XX
CS Homo sapiens.
XX
XX Key Location/Qualifiers
XX Misc-difference 152

FT /note= "Encoded by AGG in AAH46124"

PN WO200138530-A1.

XX 31-MAY-2001.

XX 22-NOV-2000; 2000WO-JP08232.

XX 24-NOV-1999; 99JP-0333479.

PR 27-APR-2000; 2000JP-0127589.

XX (TAKE) TAKEDA CHEM IND LTD.

PA Nakanishi A, Morita S;

XX WPI: 2001-355935/37

DR N-PSDB; AAH46102, AAH46124.

XX New antisense nucleotide, useful for treatment and prevention of

PT bronchial asthma and chronic obstructive pulmonary disease -

PS Claim 2; Page 76-80; 104pp; Japanese.

XX The invention relates to an antisense nucleotide targetted to the mouse

CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,

CC the CLC1 gene (coding sequence shown in AAH46102). The invention also

CC relates to an antibody specific for the Gob-5 protein, medical and

CC diagnostic compositions containing the antisense nucleotide or the

CC antibody, and methods and kits for screening for compounds which inhibit

CC the protein. Gob-5 and CLC1 are proteins expressed by goblet cells.

CC The antisense oligonucleotides and antibody are therefore useful for the

CC treatment and prevention of bronchial asthma and chronic obstructive

CC pulmonary disease. The present sequence represents human CLC1 protein.

XX SQ Sequence 914 AA;

alignment_scores:

Quality: 278.00 Length: 50

Ratio: 5.560 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-16 x AAB73716 ..

Align seg 1/1 to: AAB73716 from: 1 to: 914

2 TTTATTCTCCACAGACTCGCCAGAGACACCTAGTCTGATGAACGTC 51

|||||

865 PheileProGlnThrProGluThrProSerProAspGluThrSe 881

52 TGCTCTCTTGTCTTAATTCATATCAACAGACACCATTCCTGGCATTCACA 101

|||||

881 rAlaProCysProAsnIleHisIleAsnSerThrIleProGlyIleHisI 898

102 TTTTAAATTAATGCGAGTGTAGGAGACCTGAGTGTGCAATAGCC 151

|||||

898 LeuLeuIleMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB74733

seq_documentation_block:

ID AAB74733 standard; Protein; 914 AA.

XX AAB74733;

AC

XX 12-JUN-2001 (first entry)

DT

XX Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.

DE

XX Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;

KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;

KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;

KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;

KW anti-angiogenic; antiparkinsonian; antimicrobial; vulnery; gene therapy;

KW immune disorder; hyperproliferative disorder; cardiovascular disease;

KW cancer; angiogenic disorder; neurological disorder; infectious disease;

KW wound healing; regeneration; chemotaxis; chromosome 1.

XX OS Homo sapiens.

XX W0200112775-A2.

XX 22-FEB-2001.

XX 16-AUG-2000; 2000WO-US22325.

XX 17-AUG-1999; 99US-0149182.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Roßen CA, NI J, Florence KA, Fiscella M, Wei P, Baker KP;

PI Birse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;

XX WPI: 2001-147550/15.

DR N-PSDB; AAF81787.

XX Nucleic acids encoding 25 human secreted polypeptides, useful for

PT preventing, diagnosing and/or treating e.g. cancers, Parkinson s

PT Disease and diabetic retinopathy -

XX Claim 11; Page 459-460; 485pp; English.

XX AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733

CC to AAB74772. Human secreted proteins can have activities based on the

CC tissues and cells they are expressed in. Example of activities include:

CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;

CC antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;

CC vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic;

CC anticonvulsant; anti-angiogenic; antiparkinsonian; antimicrobial; and

CC vaccine. Human secreted protein nucleotide sequences (NAM1) and proteins

CC (PEP1) may be used in the prevention, diagnosis and treatment of diseases

CC associated with inappropriate polypeptide expression. For example, NAM1

CC and PEP1 may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patients genome

CC that affect the activity of proteins by expressing inactive proteins or

CC to supplement the patients own production of polypeptides. Disorders that

CC may be prevented, diagnosed and/or treated include immune disorders,

CC hyperproliferative disorders (e.g. cancers), cardiovascular diseases,

CC angiogenic disorders, neurological disorders, infectious diseases and/or

CC for promoting wound healing, regeneration and /or chemotaxis. AAF81778 to

CC AAF81786 and AAB74732 represent sequences used in the exemplification of

CC the present invention.

XX SQ Sequence 914 AA;

alignment_scores:

Quality: 278.00 Length: 50

Ratio: 5.560 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-16 x AAB74733 ..

Align. seg 1/1 to: AAB74733 from: 1 to: 914

2 TTTATTCTCCACAGACTCGCCAGAGACACCTAGTCTGATGAACGTC 51

|||||

865 PheileProGlnThrProGluThrProSerProAspGluThrSe 881

52 TGCTCTCTTGTCTTAATTCATATCAACAGACACCATTCCTGGCATTCACA 101

|||||

881 rAlaProCysProAsnIleHisIleAsnSerThrIleProGlyIleHisI 898

102 TTTTAAATTTATGCGAAGTAGGAGAACTGCAGCTGCTCAATAGCC 151
|||||
898 leLeuLysileMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleala 914
seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAG75474

seq_documentation_block:
ID AAG75474 standard: Protein; 925 AA.
XX
AC AAG75474;
XX
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6238.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 1.
XX
OS Homo sapiens.
XX
PN WC20012920-A2.
XX
PD 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI; 2001-235357/24.
DR N-PSDB; AAH34879.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 11; Page 7686-7690; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patients own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent sequences used in the exemplification of the present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 925 AA;

alignment_scores:
Quality: 278.00 Length: 50
Ratio: 5.560 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-16 x AAG75474 ..

Align seg 1/1 to: AAG75474 from: 1 to: 925

TTTATTCCTCCACAGACTCCGCCAGACACTAGTCTGTGATGAACGTC 51
|||||
876 PheIleProGlnThrProGlnThrProSerProAspGluThrSe 892
TGCTCCCTGTCTAATATTCATATCAACAGCACCATTCTCGCATTCACA 101
|||||
894 AlaProCysProAsnIleHisIleAsnSerThrIleProGlyIleHisI 909
TTTTAAAAATTTATCTGGAAGTGGATAGCAGAACTGCAGCTGCTCAATAGCC 151
|||||
909 leLeuLysileMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleala 925
seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74822

seq_documentation_block:
ID AAB74822 standard: Protein; 913 AA.

XX
AC AAB74822;
XX
DT 13 JUN-2001 (first entry)
XX
DE Murine ICACC-1 protein sequence.
XX
KW ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW inflammatory bowel disease; autoimmune disease.
XX
OS Mus sp.

XX
XX WO9944620-A1.
XX
XX 10-SEP-1999.
XX
XX 03-MAR-1999; 99WO-US04703.
XX
XX 03-MAR-1998; 98US-0076815.
XX
XX (MAGA-) MAGAININ PHARM INC.

XX
XX Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;
PI Nicolaidis NC, Zhou Y, Dong Q;
XX
XX WPI; 1999-350979/46.
DR N-PSDB; AAF81925.
XX

PT New nucleic acid encoding calcium activated chloride channel, used to
PT identify, e.g. specific modulators for treating atopic allergy -
XX
XX Claim 12; Fig 2; 75pp; English.

XX
XX The present sequence represents the murine interleukin 9 (IL-9) induced
XX calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
XX have anti-allergic, anti-asthmatic, anti-inflammatory and
XX immunomodulatory activities. Compounds (A) that downregulate ICACC are
XX used to alleviate asthma (or more generally atopic allergy), while those
XX (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
XX inflammatory bowel disease (IBD) (or other autoimmune diseases).
XX Measuring levels of ICACC-1 can be used in the diagnosis of asthma
XX (increased levels) or IBD (reduced levels), also for monitoring
XX treatment of these conditions. The ICACC proteins can be used:
XX (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
XX reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
XX (A) to disrupt bonding between ICACC and its ligands; and (iii) to
XX identify modulators and binding partners. ICACC polynucleotides can be
XX used to generate transgenic animals or recombinant cells, used to screen
XX for antagonists, also as a source of therapeutic antisense agents or
XX diagnostic probes (for quantifying mRNA expression, e.g. for
XX identification of modulators).

SQ Sequence 913 AA;

SQ Sequence 913 AA;

alignment_scores:
Quality: 179.50 Length: 49
Ratio: 4.378 Gaps: 1
Percent Similarity: 83.673 Percent Identity: 61.224

alignment_block:
US-09-049-696-16 x AAB73715 ..

Align seg 1/1 to: AAB73715 from: 1 to: 913

2 TTTATTCTCCACAGACTCGCCGAGACACCTAGTCTCATGAACGTC 51
||||| ||||| |||||
865 PheileProAlaGlnGlupro.....ProlleProGluasperTh 878

52 TGCTCCTTGCTCTAATATCATATCAACAGACACATTCCTGGCATTCACA 101
: |||||::: ||| ||||| ||||| ||||| |||||
878 rProProcysProAspileSerilleAsnSerThrIleProGlyIleHisV 895

102 TTTTAAAAATATCTGGAAGTGGATAGAGAAGCTGCAGCTGTCAATA 148
: :|||::: ||||| ||||| ||||| ||||| |||||
895 alLeuLysileMetTrpLysrpleGlyGluMetGlnValThrLeu 910

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.AAW06548

seq_documentation_block:

ID AAW06548 standard; Protein; 228 AA.

XX AC AAW06548;

XX DT 13-MAR-1997 (first entry)

XX DE Human colon specific gene CSG5 polypeptide fragment.

XX KW Colon specific gene; CSG5; colon cancer; metastasis; diagnosis;

XX FW therapy; antibody; vaccine.

XX OS Homo sapiens.

XX PN W09639419-A1.

XX FD 12-DEC-1996.

XX PF 06-JUN-1995; 95WO-US07289.

XX PR 06-JUN-1995; 95WO-US07289.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosén CA, Yu G;

XX DR WP.; 1997-043054/04.

XX DR N:P5DB; AAT45884.

XX PT Human colon specific genes and their expression products - detection of which, in non-colon tissue samples, can be used as indication of colon cancer metastasis

XX PS Claim 8; Fig 5; 60pp; English.

XX NCxvcl polypeptides (AAW06545-53) are encoded by cDNA clones (see also AAU45880-92) corresponding to 13 human colon specific genes, designated CSG1, CSG2, etc., that are primarily expressed in tissues derived from the colon. Recombinant CSG polypeptides can be produced in transformed host cells. They are useful diagnostic markers for colon cancer and for colon cancer metastasis and can also be used to screen for (antagonist cpds. of therapeutic or diagnostic value. Antibodies raised against the colon-specific polypeptides may be used to target colon cancer cells or as part of a colon cancer vaccine.

XX CC

SQ Sequence 228 AA;

alignment_scores:
Quality: 116.00 Length: 20
Ratio: 5.800 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-16 x AAW06548 ..

Align seg 1/1 to: AAW06548 from: 1 to: 228

2 TTTATTCTCCACAGACTCCGCCAGACACCTAGTCTGTGTAAGAGTC 51
|||||
209 PheileProGlnThrProProGluThrProSerProAspGluThrSe 225
52 TGCTCTCTGT 61
|||||
225 rAlaProCys 228

seq_name: /SDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW46879

seq_documentation_block:
ID AAW46879 standard; Protein; 228 AA.

XX AC AAW46879;
XX DT 22-JUN-1998 (first entry)
XX OS Homo sapiens.
XX KW Colon-specific gene; probe; detection; expression; human;
XX DE diagnostic assay; colon cancer; antibody; screening.
XX OS Homo sapiens.
XX PN US5733748-A.
XX PD 31-MAR-1998.
XX PF 06-JUN-1995; 95US-0469667.
XX PR 06-JUN-1995; 95US-0469667.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen C, Yu G;
XX WPI: 1998-229823/20.
XX N-PSDB; AAV16672.

Colon-specific nucleic acids - useful as probes for detecting colon cancer micrometastases

Claim 1; Fig 5; 51pp; English.

AAW46876-80 and W4682-85 represent proteins encoded by colon-specific genes. The polynucleotides encoding these proteins can be used as probes to detect expression of the corresponding human genes, e.g. in diagnostic assays for detecting micrometastases of colon cancer. Recombinant cells containing the polynucleotides can be used to produce the proteins, in order that antibodies can be raised and used in further screening or diagnostics.

SQ Sequence 228 AA;

alignment_scores:
Quality: 116.00 Length: 20
Ratio: 5.800 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-16 x AAW46879 ..

Align seg 1/1 to: AAW46879 from: 1 to: 228

12 TTTATTCTCCACAGACTCCGCCAGACACCTAGTCTGTGTAAGAGTC 51
|||||
209 PheileProGlnThrProProGluThrProSerProAspGluThrSe 225
52 TGCTCTCTGT 61
|||||
225 rAlaProCys 228

seq_name: /SDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA892738

seq_documentation_block:
ID AAB92738 standard; Protein; 339 AA.

XX AC AAB92738;
XX DT 26-JUN-2001 (first entry)
XX DE Human protein sequence SEQ ID NO:11180.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.
XX PN EF1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX RA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 8; SEQ ID 11180; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 339 AA;

alignment_scores: Length: 35
 Quality: 66.00
 Ratio: 3.300
Percent Similarity: 57.143 Percent Identity: 40.000

alignment_block:

US-09-049-696-16 x AAB92738 ..
Align seg 1/1 to: AAB92738 from: 1 to: 339
8 CCTCCACAGACTCCGCCAGAGACACCTAGTCTGTGATGAACGTCGTCTCC 57
||||| ||||| ||| : : :
186 ProProGlnProProGlnProGlnProProProProSerProSerProSer 202
58 TTGTCTCTAAT.....ATTCTATATCAACAGACACCATTC 89
|: || : : : :
202 OAlaArgAsnProProGlyAlaSerThrTyraSnLysAsnSerAsnIleP 219
90 CTGGC 94
|||||
219 roGly 220

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB94896

seq_documentation_block:

ID AAB94896 standard; Protein: 390 AA.
XX
AC AAB94896;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:16244.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -

PS Claim 8; SEQ ID 16244; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides, and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 390 AA;

alignment_scores:

 Quality: 66.00 Length: 35
 Ratio: 3.300 Gaps: 1
Percent Similarity: 57.143 Percent Identity: 40.000

alignment_block:

US-09-049-696-16 x AAB94896 ..
Align seg 1/1 to: AAB94896 from: 1 to: 390
3 CCTCCACAGACTCCGCCAGAGACACCTAGTCTGTGATGAACGTCGTCTCC 57
||||| ||||| ||| : : :
227 ProProGlnProProGlnProProProProProProSerProSerProSer 243
58 TTGTCTCTAAT.....ATTCTATATCAACAGACCATTC 89
|: || : : : :
243 OAlaArgAsnProProGlyAlaSerThrTyraSnLysAsnSerAsnIleP 260
90 CTGGC 94
|||||
269 roGly 261

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB95249

seq_documentation_block:

ID AA395249 standard; Protein: 756 AA.
XX
AC AA395249;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:17413.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -

PS Claim 8; SEQ ID 16244; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

us-09-049-696-16.rag

Tue Apr 2 09:39:47 2002

90 CTGGC 94
|||||
736 roGly 737

OM of: US-09-049-696-15 to: SPTREMBL_17:* out_format : pfs

Date: Mar 30, 2002 2:47 AM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09049696/runat_28032002.145238.2053/app_query.fasta_1.12579
-DB=SPTREMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -GAPEXT=0.500
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1.705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-15
Query length: 154
Database: SPTREMBL_17*
Database sequences: 473505
Database length: 146272329
Search time (sec): 805.760000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
sp_human:O95151	+	265.00	614.37	6.6e-27	914	! O95151 homo sapiens (human). ca
sp_human:Q9UPC6	+	265.00	614.37	6.6e-27	914	! Q9UPC6 homo sapiens (human). ca
sp_human:Q9UNF6	+	265.00	614.37	6.6e-27	914	! Q9UNF6 homo sapiens (human). ca
sp_mammal:Q9TUB5	+	184.50	420.78	4.0e-16	917	! Q9TUB5 sus scrofa (pig). epiche
sp_rodent:Q88826	+	168.50	382.34	5.5e-14	913	! Q88826 mus musculus (mouse). gc
sp_rodent:Q9D726	+	168.50	382.34	5.5e-14	913	! Q9D726 mus musculus (mouse). ch
sp_human:Q9NXP1	+	118.50	267.06	2.8e-07	469	! Q9NXP1 homo sapiens (human). cd
sp_human:Q9UNF7	+	118.50	262.09	2.8e-07	917	! Q9UNF7 homo sapiens (human). ca
sp_human:Q9UQC9	+	92.50	199.36	0.0008	943	! Q9UQC9 homo sapiens (human). ch
sp_human:Q9V6N2	+	92.50	199.36	0.0008	943	! Q9V6N2 homo sapiens (human). ca
sp_plant:Q22017	+	72.00	151.48	0.4687	780	! Q22017 cylindrothea fusiformis
sp_rodent:Q9R070	+	69.00	143.19	1.17	902	! Q9R070 mus musculus (mouse). ca
sp_rodent:Q9EQR4	+	69.00	143.19	1.17	902	! Q9EQR4 mus musculus (mouse). en
sp_human:Q9BZK6	+	67.00	139.98	2.20	727	! Q9BZK6 homo sapiens (human). p8
sp_bacteria:Q67406	+	67.00	139.98	2.19	759	! Q67406 aquifex aeolicus. cellul
sp_mammal:O18741	+	66.00	135.95	2.96	905	! O18741 bos taurus (bovine). lu
sp_invertebrate:Q9VZC2	+	65.50	140.00	3.57	446	! Q9VZC2 drosophila melanogaste
sp_rodent:Q921V9	+	65.50	138.85	3.55	521	! Q921V9 mus musculus (mouse). wc
sp_rodent:Q9ES20	+	65.00	136.22	4.10	631	! Q9ES20 rattus norvegicus (rat).
sp_invertebrate:Q45360	+	65.00	133.45	4.03	916	! Q45360 caenorhabditis elegans
sp_vertebrate:Q91295	+	64.50	140.95	4.97	284	! Q91295 rana catesbeiana (bull f
sp_bacteria:Q9JYD1	+	64.00	136.39	5.67	446	! Q9JYD1 neisseria meningitidis
sp_virus:Q9WPM2	+	64.00	136.36	5.67	448	! Q9WPM2 macaca mulatta rhadinovi
sp_virus:Q9J2H0	+	64.00	136.36	5.67	448	! Q9J2H0 macaca mulatta rhadinovi
sp_bacteria:Q9JTB4	+	64.00	136.10	5.66	464	! Q9JTB4 neisseria meningitidis
sp_plant:Q9M1B1	+	64.00	134.85	5.62	549	! Q9M1B1 arabidopsis thaliana (mc
sp_plant:Q9C660	+	64.00	132.44	5.53	760	! Q9C660 arabidopsis thaliana (mc
sp_plant:Q9SFF7	+	63.50	144.58	7.03	126	! Q9SFF7 arabidopsis thaliana (mc
sp_rodent:Q60749	+	63.50	135.24	6.62	443	! Q60749 mus musculus (mouse). p8
sp_rodent:Q60735	+	63.50	135.24	6.62	443	! Q60735 mus musculus (mouse). si
sp_rodent:Q99M33	+	63.50	135.24	6.62	443	! Q99M33 mus musculus (mouse). si
sp_rodent:Q9WUN1	+	63.50	128.09	6.32	1160	! Q9WUN1 mus musculus (mouse). t
sp_rodent:Q99PM6	+	63.50	127.86	6.31	1198	! Q99PM6 mus musculus (mouse). d
sp_rodent:Q99PB3	+	63.50	124.10	6.16	1987	! Q99PB3 mus musculus (mouse). m
sp_plant:Q9LW52	+	63.00	133.89	7.72	452	! Q9LW52 arabidopsis thaliana (mc
sp_virus:Q83344	+	63.00	133.40	7.69	483	! Q83344 murine herpesvirus 72. m
sp_virus:Q9DYE3	+	63.00	133.40	7.69	483	! Q9DYE3 murine herpesvirus 72. m
sp_invertebrate:Q22354	+	63.00	129.10	7.48	862	! Q22354 caenorhabditis elegans
sp_invertebrate:Q17646	+	62.50	125.56	8.60	1181	! Q17646 caenorhabditis elegans
sp_invertebrate:Q27045	+	62.00	135.12	10.76	277	! Q27045 telleria parva. schiz

sp_bacteria:Q55106 - 62.00 135.09 10.75 278 ! Q55106 synechococcus sp. (st
sp_invertebrate:Q9V728 + 62.00 133.80 10.67 331 ! Q9V728 drosophila melanoga
sp_bacteria:Q9F6X7 + 62.00 131.65 10.52 442 ! Q9F6X7 chloroflexus aurantia
sp_bacteria:Q9AKJ0 + 62.00 130.83 10.46 494 ! Q9AKJ0 rickettsia rickettsii
sp_plant:Q65672 + 62.00 128.52 10.31 674 ! Q65672 arabidopsis thaliana

seq_name: sp_human:O95151

seq_documentation_block:

ID O95151 PRELIMINARY; PRT; 914 AA.

AC O95151;

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.

GN HCLCAL.

QS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=SMALL INTESTINE;

RX MEDLINE=99047526; PubMed=9828122;

RA Gruber A.D., Elbie R.C., Ji H.L., Schreier K.D., Fuller C.M.,

RA Pauli B.U.;

RT "Genomic cloning, molecular characterization, and functional analysis

RT of human CLCAL, the first human member of the family of Ca2+-activated

RL Genomics 54:200-214(1998).

DR EMBL: AF039400; AAC95428.1;

DR InterPro: IPR000131; ATase_gamma.

DR InterPro: IPR002035; VWFA.

DR PROSITE: P500153; ATPASE_GAMMA; UNKNOWN_1.

DR PROSITE: P500234; VWFA; 1.

DR SMART: SM00327; VWA; 1.

SQ SEQUENCE 914 AA; 100174 MW; 7066440C46557FA3 CRC64;

alignment_scores:

Quality: 265.00 Length: 51

Ratio: 5.196 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-15 x O95151

Align seg 1/1 to: O95151 from: 1 to: 914

1 ACAGATCTTTTCATTGCTATTTCAGCTGTTGATGAAGTGCATCTGAATC 50

|||||

838 ThrAspLeuPheLeAlaLeGlnAlaValAspLysValAspLeuLysSe 854

51 AGAATAATCAACATTCACGAGATCTTTCTTTTATTCCTCCACAGACTC 100

|||||

854 rGluLeuSerAsnLeuAlaArgValSerLeuPheLeuProGlnThrP 871

101 CGCAGACAGACCTAGTCCTGATGAACAGCTGCTCTTCCTTCCTAATATT 150

|||||

871 roProGluThrProSerProAspGluThrSerAlaProCysProAsnIle 887

151 CAT 153

|||||

888 His 888

seq_name: sp_human:Q9UPC6

seq_documentation_block:

ID Q9UPC6 PRELIMINARY; PRT; 914 AA.

AC Q9UPC6;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

```

DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Elble R.C., Ji H.L., Schreuer K.D., Fuller C.M.,
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
RT of human CLCAL, the first human member of the family of Ca2+-activated
RT Cl-channel proteins."
RL Genomics 54:200-214(1998).
DR EMBL; AF039401; AAC95429.1; -.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWFA; 1.
DR SEQUENCE 914 AA; 100160 MW; 7761362B66752AA3 CRC64;
SQ

alignment_scores:
  Quality: 265.00      Length: 51
  Ratio: 5.196        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-15 x Q9UPC6 ..
Align seg 1/1 to: Q9UPC6 from: 1 to: 914

1 ACAGATCTTTTCATTCCTATTCAGGCTGTGATAGGTCGATCGAATC 50
|||||
838 ThrAspLeuPheIleAlaIleGlnAlaValSerLeuPheIleProGlnThrP 854

51 AGAAATATCCAACTGCAGAGTATCTTTGTTTATCTCCACAGACTC 100
|||||
854 rGluIleSerAsnIleAlaArgValSerLeuPheIleProGlnThrP 871

101 CGCCAGACACCTAGTCTGATGAAGCTGCTGCTGCTGCTGCTAATATT 150
|||||
871 roProGluThrProSerProAspGluThrSerAlaProCysProAsnIle 887

151 CAT 153
|||
888 His 888

seq_name: sp_human.Q9UNF6

seq_documentation_block:
ID Q9UNF6 PRELIMINARY; PRT; 914 AA.
AC Q9UNF6;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 1.
GN CACCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=SMALL INTESTINE, AND COLON;
RX MEDLINE=99364503; PubMed=10437792;
RA Agnel M., Vermet T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
RT chloride channel (cacC) family predominantly expressed in the
RT digestive tract and trachea."
RL FEBS Lett. 455:295-301(1999).

```

```

DR EMBL; AF127036; AAD25487.1; -.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
DR SEQUENCE 914 AA; 100202 MW; B58B451D97887D6B CRC64;
SQ

alignment_scores:
  Quality: 265.00      Length: 51
  Ratio: 5.196        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-15 x Q9UNF6 ..
Align seg 1/1 to: Q9UNF6 from: 1 to: 914

1 ACAGATCTTTTCATTCCTATTCAGGCTGTGATAGGTCGATCGAATC 50
|||||
838 ThrAspLeuPheIleAlaIleGlnAlaValSerLeuPheIleProGlnThrP 854

51 AGAAATATCCAACTGCAGAGTATCTTTGTTTATCTCCACAGACTC 100
|||||
854 rGluIleSerAsnIleAlaArgValSerLeuPheIleProGlnThrP 871

101 CGCCAGACACCTAGTCTGATGAAGCTGCTGCTGCTGCTGCTAATATT 150
|||||
871 roProGluThrProSerProAspGluThrSerAlaProCysProAsnIle 887

151 CAT 153
|||
888 His 888

seq_name: sp_mammal.Q9TUB5

seq_documentation_block:
ID Q9TUB5 PRELIMINARY; PRT; 917 AA.
AC Q9TUB5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN.
GN ABCC.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9623;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=ILEAL MUCOSA;
RX Gaspar K.J., Gabriel S.E., Racette K.J., Forsyth G.W.;
RX "A cDNA involved in porcine exocrine chloride conductance."
RX Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF095584; AAF00077.1; -.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWFA; 1.
DR SEQUENCE 917 AA; 100735 MW; 649F41944F7EAD19 CRC64;
SQ

alignment_scores:
  Quality: 184.50      Length: 54
  Ratio: 4.011        Gaps: 2
Percent Similarity: 85.185 Percent Identity: 70.370

alignment_block:
US-09-049-696-15 x Q9TUB5 ..
Align seg 1/1 to: Q9TUB5 from: 1 to: 917

```

1 ACAGATCTTTTCATGCTATTGCTAGGCTGTTGATTAAGTTCGATCTGAATC 50
|||||
840 ThrAspLeuPheIleAlaValGlnAlaValAspLysThrAsnLeuLysSe 856
51 AGAATATCCAACTTCACGAGTATCTTTCTTTATT
|||||
856 rGluLeuSerAsnIleAlaGlnValSerLeuPheLeuProGluAlaLap 873
89 CTCACAGACTCCGCCAGAGACACTAGTCTGATGAACACTGCTCCTCT 138
|||||
873 roProGluThrProGluThrProAlaPro.....SerLeuPro 886
139 TGTCTTAATATT 150
|||||
887 CysProGluIle 890
seq_name: sp_rodent:088826
seq_documentation_block:
ID 088826; PRELIMINARY; PRT; 913 AA.
AC 088826;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE GDB-5 PROTEIN
GN CLCA3 OR GDB-5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RA Komiya T., Tanigawa Y., Hirohashi S.;
RT "Cloning of a gene, gob-5, which is expressed in intestinal goblet
cells in mice."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AH017156; BAA33743.1; -
DR MGD; MGI:1346342; Clca3.
DR InterPro: IPR000131; ATPase_gamma.
DR SMART: SM00327; VWFA; 1.
DR InterPro: IPR002035; VWFA.
DR PROSITE: PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE: PS0234; VWFA; 1.
SQ SEQUENCE 913 AA; 100070 MW; A7FA2F9E1089806D CRC64;

alignment_scores:
Quality: 168.50 Length: 50
Ratio: 4.012 Gaps: 1
Percent Similarity: 84.000 Percent Identity: 68.000

alignment_block:
US-09-049-696-15 x 088826
Align seg 1/1 to: 088826 from: 1 to: 913
1 ACAGATCTTTTCATGCTATTGCTAGGCTGTTGATTAAGTTCGATCTGAATC 50
|||||
838 ThrAspLeuPheIleAlaValGlnAlaValAspLysSerAsnLeuLysSe 854
51 AGAATATCCAACTTCACGAGTATCTTTCTTTATT
|||||
854 rGluLeuSerAsnIleAlaGlnValSerValPheIleProAlaGlnLap 871
101 CGCCAGAGACACTAGTCTGCTGATGAACGCTGCTCTCTCTTAATATT 150
|||||
871 ro.....ProIleProGluAspSerThrProProCysProAspIle 884
seq_name: sp_rodent:090726
seq_documentation_block:

ID 090726; PRELIMINARY; PRT; 469 AA.
AC 090726;
DT 01-OCT-2000 (TReMBLrel. 15, Created)

Q9D726; PRELIMINARY; PRT; 913 AA.
Q9D726;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE GLORIIDE CHANNEL CALCIUM ACTIVATED 3.
GN CLCA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=STOMACH;
MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK008659; BAB25815.1; -
DR MGD; MGI:1346342; Clca3.
DR InterPro: IPR000131; ATPase_gamma.
DR SMART: SM00327; VWFA; 1.
DR PROSITE: PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE: PS0234; VWFA; 1.
SQ SEQUENCE 913 AA; 100111 MW; 491E8584B06D9A89 CRC64;

alignment_scores:
Quality: 168.50 Length: 50
Ratio: 4.012 Gaps: 1
Percent Similarity: 84.000 Percent Identity: 68.000

alignment_block:
US-09-049-696-15 x Q9D726
Align seg 1/1 to: Q9D726 from: 1 to: 913

1 ACAGATCTTTTCATGCTATTGCTAGGCTGTTGATTAAGTTCGATCTGAATC 50
|||||
838 ThrAspLeuPheIleAlaValGlnAlaValAspLysSerAsnLeuLysSe 854
51 AGAATATCCAACTTCACGAGTATCTTTCTTTATT
|||||
854 rGluLeuSerAsnIleAlaGlnValSerValPheIleProAlaGlnLap 871
101 CGCCAGAGACACTAGTCTGCTGATGAACGCTGCTCTCTCTTAATATT 150
|||||
871 ro.....ProIleProGluAspSerThrProProCysProAspIle 884
seq_name: sp_human:Q9NXP1
seq_documentation_block:

ID Q9NXP1; PRELIMINARY; PRT; 469 AA.
AC Q9NXP1;
DT 01-OCT-2000 (TReMBLrel. 15, Created)

```

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE CDNA FLJ20131 FIS, CLONE COL06357.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Oabayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000138; BAA90969.1; -.
SQ SEQUENCE 469 AA; 50694 MW; 6765986C886B5E356 CRC64;

alignment_scores:
  Quality: 118.50      Length: 57
  Ratio: 3.118        Gaps: 1
  Percent Similarity: 66.667      Percent Identity: 45.614

alignment_block:
US-09-049-696-15 x Q9UNF7 ..
Align seg 1/1 to: Q9UNF7 from: 1 to: 469

1 ACAGATCTTTTCATTGCTATTAGGCTGTGTATAGGTCGATCGAATC 50
||| ::::::::::::::::::::::::::::::::::::: |||
391 ThrHisilePheileAlaileLysSerileAspLysSerAsnLeuThrSe 407
| :::::::::::::::::::::::::::::::::::::::
51 AGAATATCCACATTCGACGAGTATCTTTGTTTATTCCT ..... 90
| :::::::::::::::::::::::::::::::::::::::
407 rLysValSerAsnileAlaileGlnValThrLeuPheileProGlnAlaAsnP 424
91 .....CCACAGACTCCGCCAGACCTAGTCTGTGATGAACG 129
424 roAspAspIleAspProThrProThrProThrProThrProAspLysSer 440
130 TCTGCTCTTGTCTCTATATT 150
441 HisAsnSerGlyValAsnile 447

seq_name: sp_human:Q9UNF7

seq_documentation_block:
ID Q9UNF7 PRELIMINARY; PRT; 917 AA.
AC Q9UNF7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 2.
GN CACCC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RX MEDLINE=99364503; PubMed=10437792;
RA Agnel M., Vermet T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
RT chloride channel (CaCC) family predominantly expressed in the
RT digestive tract and trachea.";
RL PBBS Lett. 455:295-301(1999).
DR EMBL; AF127035; AAD48398.1; -.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00092; vwa; 1.
DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWA; 1.

```

```

SQ SEQUENCE 917 AA; 101153 MW; EA01090D781BEB95 CRC64;

alignment_scores:
  Quality: 118.50      Length: 57
  Ratio: 3.118        Gaps: 1
  Percent Similarity: 66.667      Percent Identity: 45.614

alignment_block:
US-09-049-696-15 x Q9UNF7 ..
Align seg 1/1 to: Q9UNF7 from: 1 to: 917

1 ACAGATCTTTTCATTGCTATTAGGCTGTGTATAGGTCGATCGAATC 50
||| ::::::::::::::::::::::::::::::::::::::: |||
839 ThrHisilePheileAlaileLysSerileAspLysSerAsnLeuThrSe 855
51 AGAATATCCACATTCGACGAGTATCTTTGTTTATTCCT ..... 90
| :::::::::::::::::::::::::::::::::::::::
855 rLysValSerAsnileAlaileGlnValThrLeuPheileProGlnAlaAsnP 872
91 .....CCACAGACTCCGCCAGACCTAGTCTGTGATGAACG 129
872 roAspAspIleAspProThrProThrProThrProThrProAspLysSer 888
130 TCTGCTCTTGTCTCTATATT 150
883 HisAsnSerGlyValAsnile 895

seq_name: sp_human:Q9UNF7

seq_documentation_block:
ID Q9UNF7 PRELIMINARY; PRT; 943 AA.
AC Q9UNF7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CHLORIDE CHANNEL PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CORNEAL EPITHELIUM;
RA Itah R., Kawamoto S., Kinoshita S., Kawasaki S., Okubo K.;
RT "Isolation and characterization of chloride channel in human corneal
RL epithelium.";
RL EMBL; AB026833; BAA77810.1; -.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 943 AA; 103939 MW; 4059766CFBD20E72 CRC64;

alignment_scores:
  Quality: 92.50      Length: 39
  Ratio: 2.803        Gaps: 1
  Percent Similarity: 84.615      Percent Identity: 46.154

alignment_block:
US-09-049-696-15 x Q9UNF7 ..
Align seg 1/1 to: Q9UNF7 from: 1 to: 943

7 CTTTCATTGCTATTAGGCTGTGTATAGGTCGATCGAATCAGAAAT 56
:::::::::::::::::::::::::::::::::::::::::::::
862 IleTyrrValAlaileAlaileArgAlaileAspArgAsnSerLeuGlnSerAlaVa 878
57 ATCCAACATTCGACGAGTATCTTTGTTTATTCCTCCACAGACTCCGCCAG 106
| ::::::::::::::::::::::::::::::::::::::: |||
878 lSerAsnileAlaileGlnAlaProLeuPheileProProAsnSerAspPro. 894

```

107 AGACACCTAGTCTCGAT 123
:::||||: |||
895 ..ValProAlaArgAsp 899

seq_name: sp_human:Q9Y6N2

seq_documentation_block:
ID Q9Y6N2 PRELIMINARY; PRT; 943 AA.

AC Q9Y6N2;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL-2.
GN hCLCA2 OR CACCC3.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99292364; PubMed=10362588;
RA Gruber A.D., Schreier K.D., Ji H.-L., Fuller C.M., Pauli B.U.;
RT "Molecular cloning and transmembrane structure of hCLCA2 from human
lung, trachea, and mammary gland.";
RT lung, trachea, and mammary gland.";
RL Am. J. Physiol. 276:C1261-C1270(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99364503; PubMed=10437792;
RA Agnel M., Vermet T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
chloride channel (CaCC) family predominantly expressed in the
digestive tract and trachea.";
RL FEBS Lett. 455:295-301(1999).
DR EMBL; AF043977; AAD40367.1; -
DR EMBL; AF127980; AAD48397.1; -
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS0234; VWFA; 1.
SQ SEQUENCE 943 AA; 103940 MW; 0E09A0909D2529B CRC64;

alignment_scores:
Quality: 92.50 Length: 39
Ratio: 2.803 Gaps: 1
Percent Similarity: 84.615 Percent Identity: 46.154

alignment_block:

US-09-049-696-15 x Q9Y6N2 ..

Align seg 1/1 to: Q9Y6N2 from: 1 to: 943

7 CTTTTCATTGCTATTCAGGCTGTGATAAGCTGCTGAAATCAGAAAT 56
:::||||: |||
862 IletyrValAlaIleArgAlaMetAspArgAsnSerLeuGlnSerAlaVa 878
57 ATCCAACTGCAGAGATATCTTTGTTATTCCTCCACAGACTCCGCCAG 106
:|||||: |||
878 lSerAsnIleAlaGlnAlaProLeuPheIleProProAsnSerAspPro. 894

107 AGACACCTAGTCTCGAT 123
:::||||: |||
895 ..ValProAlaArgAsp 899

seq_name: sp_plant:O22017

seq_documentation_block:

ID O22017 PRELIMINARY; PRT; 780 AA.

AC O22017;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE HEPC PROTEIN.

GN HEPC.
OS Cyllindrotheca fusiformis (Marine diatom).
OC Eukaryota; stramenopiles; Bacillariophyta; Bacillariophyceae;
OC Bacillariophycidae; Bacillariales; Bacillariaceae; Cyllindrotheca.
OX NCBI_TaxID=2853;
RN [1]
RP SEQUENCE FROM N.A.
RA Kroeber N., Lehmann G., Rachel R., Sumner M.;
RL Eur. J. Biochem. 0:0-0(0).
DR EMBL; Y15086; CAA75362.1; -
DR Mendel; 23994; Cylfu:3111; 23994.
SQ SEQUENCE 780 AA; 83687 MW; 2661BC277949F691 CRC64;

alignment_scores:
Quality: 72.00 Length: 59
Ratio: 2.323 Gaps: 3
Percent Similarity: 52.542 Percent Identity: 35.593

alignment_block:

US-09-049-696-15 x O22017 ..

Align seg 1/1 to: O22017 from: 1 to: 780

7 CTTTTCATTGCTATTCAGGCTGTGATAAGCTC..... 39
:||||: |||
9 LeuPheLeuThrAlaGlnGlyAlaThrSerValAlaThrAspAlaSerG1 25
40GATCTGAATCAGAA.....ATATCCAACATTGCAC 70
|||||: |||
25 uGlnAspLeuLysGlnGluAspGluAlaTrpTrpSerArgMetValGlnG 42
71 GAGTATCTTTGTTTATTCCTCCACAGACTCCGACAGACACCTAGTCT 120
:||||: |||
42 luValAsnSerLeuThrProProProProProProProProProProPro 58
121 GATCAACAGCTGCTCTCTCTGCTTAAT 147
||| |||
59ThrArgProProThrProAsn 65

seq_name: sp_rodent:Q9R070

seq_documentation_block:

ID Q9R070 PRELIMINARY; PRT; 902 AA.

AC Q9R070;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE CA(2+)-SENSITIVE CHLORIDE CHANNEL 2.
GN CLCA2 OR CACC.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY;
RX MEDLINE=20012773; PubMed=10544033;
RA Lee D., Ha S., Kho Y., Kim J., Cho K., Baik M., Choi Y.;
RT "Induction of mouse Ca(2+)-sensitive chloride channel 2 gene during
involution of mammary gland.";
RL Biochem. Biophys. Res. Commun. 264:933-937(1999).
DR EMBL; AF108501; AAF12731.1; -
DR MG5; MGI:1931471; Clca2.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS0234; VWFA; 1.
SQ SEQUENCE 902 AA; 99866 MW; 0FB746CF5FB3D07F CRC64;

alignment_scores:
Quality: 69.00 Length: 28
Ratio: 3.000 Gaps: 0

Align seg 1/1 to: 067406 from: 1 to: 759

```
7 CTTTTCATTGCTATTTCAGGCTGTTGATAAGGTCGATCTGAAATCAGAAAT 56
  ::|||:::|||||::: ::|||::: ::||| ::|||
125 IlePhePheSerIleArgProIleAspArgLysAlaIleLysValAspMe 141
  ::|||::: CAGACTCCGC 103
57 ATCCAAACATTGCACGAGTATCTTTGTTTATTCCTCCA...CAGACTCCGC 103
  ::|||::: |||:::|||||::: |||
141 tGluSerLeuProThrValAspValPheIleProThrTyrAsnGluProp 158
  ||||| ||| |||||:::||||
104 CAGAGACACCTAGTCTGATGAAACGTCGTGCT 135
  ||||| ||| |||||:::||||
158 roGluIlePro.....GluThrThrAla 165
```

This Page Blank (uspto)

OM of: US-09-049-696-15 to: SwissProt_39:* out_format : pfs

Date: Mar 30, 2002 2:52 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+np2.model -DEV=xlp
-Q/cgn2_1/USPTO_spo1/US09049696/runat_28032002_145238_2085/app_query.fasta_1.12579
-DB=SwissProt_39 -QFMT=fastan -SUFFIX=isp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1_165 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-15

Query length: 154

Database: SwissProt_39:*

Database sequences: 100059

Database length: 36564827

Search time (sec): 306.030000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
SwissProt_39:ECIC_BOVIN	+	83.50	166.30	0.0152	903 P54281 bos taurus (bovine).
SwissProt_39:SUIS_RAT	+	64.50	118.27	3.52	1840 P23739 rattus norvegicus (rat)
SwissProt_39:XRCI_MOUSE	+	62.00	121.18	7.08	631 Q60596 mus musculus (mouse).
SwissProt_39:RCHD_HELMO	+	62.00	120.75	7.09	666 Q92962 heliobacillus mobilis
SwissProt_39:NPL2_MOUSE	+	61.50	122.56	8.13	460 P51860 mus musculus (mouse).
SwissProt_39:GYS_DROME	+	60.50	117.13	10.89	689 Q9VFC8 drosophila melanogaster
SwissProt_39:FIMC_SALTY	+	60.00	124.71	12.34	230 P37923 salmonella typhimurium
SwissProt_39:CDNC_MOUSE	+	60.00	121.43	12.43	348 P49919 mus musculus (mouse).
SwissProt_39:ELA_ADRO7	+	60.00	118.04	12.51	534 P40602 arabidopsis thaliana
SwissProt_39:APG_MOUSE	+	59.50	122.59	14.27	261 P03256 human adenovirus type
SwissProt_39:DIAL_HUMAN	+	59.50	110.20	14.64	1248 Q06010 homo sapiens (human)
SwissProt_39:PGCA_CHICK	+	59.50	106.04	14.76	2109 P07898 gallus gallus (chicken)
SwissProt_39:NUSM_CHICK	+	58.50	113.70	19.26	605 P18940 gallus gallus (chicken)
SwissProt_39:YAR2_SCHPO	+	58.50	107.81	19.49	1273 Q10135 schizosaccharomyces
SwissProt_39:HXAX_MOUSE	+	58.00	115.05	22.11	443 P02831 mus musculus (mouse).
SwissProt_39:SPG7_DICDI	+	58.00	113.60	22.17	532 P22698 dictyostelium discoide
SwissProt_39:MP3_HUMAN	+	58.00	112.85	22.21	585 Q13368 homo sapiens (human)
SwissProt_39:MYOD_RAT	+	57.50	116.57	25.37	318 Q02346 rattus norvegicus (rat)
SwissProt_39:YY06_CAEEL	+	57.50	113.12	25.55	491 P48460 caenorhabditis elegans
SwissProt_39:NRG3_HUMAN	+	57.50	110.09	25.71	720 P56975 homo sapiens (human).
SwissProt_39:HSF1_HUMAN	+	57.00	111.42	29.52	529 Q00613 homo sapiens (human).
SwissProt_39:CO9_RAT	+	57.00	111.05	29.54	554 Q62930 rattus norvegicus (rat).
SwissProt_39:SLP2_CLOTH	+	57.00	109.34	29.64	688 Q06853 clostridium thermocel
SwissProt_39:YDBE_SCHPO	+	57.00	107.58	29.75	859 Q10362 schizosaccharomyces
SwissProt_39:PERT_HUMAN	+	57.00	106.92	29.79	933 P07202 homo sapiens (human).
SwissProt_39:GAT2_CHICK	+	56.50	111.31	33.99	466 P23824 gallus gallus (chicken)
SwissProt_39:PSAD_HORVU	+	56.00	116.70	38.70	205 P36213 hordeum vulgare (barl
SwissProt_39:VE4_HPVO5	+	56.00	115.29	38.81	245 P06924 human papillomavirus b
SwissProt_39:VE4_HPVS8	+	56.00	115.29	38.81	245 P26550 human papillomavirus d
SwissProt_39:CP74_LINUS	+	56.00	109.08	39.30	536 P48417 linum usitatissimum
SwissProt_39:BRDI_SCHPO	+	56.00	109.00	39.31	542 Q09948 schizosaccharomyces
SwissProt_39:ODP2_HUMAN	+	56.00	108.01	39.39	614 P10515 h drosophila poamide ad
SwissProt_39:GUN6_DICDI	+	56.00	106.91	39.48	705 P22699 dictyostelium discoide
SwissProt_39:DA_DROME	+	56.00	106.86	39.48	710 P11420 drosophila melanogaster
SwissProt_39:UTX_MOUSE	+	56.00	101.87	39.88	1333 Q70546 mus musculus (mouse).
SwissProt_39:UTX_HUMAN	+	56.00	101.47	39.92	1401 Q15550 homo sapiens (human).
SwissProt_39:RB12_RAT	+	55.50	115.90	44.62	197 P35284 rattus norvegicus (rat)
SwissProt_39:ELK4_HUMAN	+	55.50	109.70	45.19	431 P28324 homo sapiens (human).
SwissProt_39:SNP_HUMAN	+	55.50	103.70	45.74	919 Q9Y6H5 homo sapiens (human).
SwissProt_39:NO5_PEA	+	55.00	117.78	51.17	135 P25226 pisum sativum (garden pe

alignment_scores:

Quality: 83.50

Length: 40

SwissProt_39:YWIS_WHEAT + 55.00 116.89 51.26 151 | P14723 triticum aestivum
SwissProt_39:Y091_NPVOP + 55.00 112.03 51.77 279 | O10341 orgyia pseudotsuga
SwissProt_39:IORI_HUMAN + 55.00 111.04 51.88 316 | P23490 homo sapiens (huma
SwissProt_39:LPAX_HUMAN + 55.00 109.45 52.05 386 | O60711 homo sapiens (huma
SwissProt_39:CGBI_CRILO + 55.00 108.62 52.13 429 | Q08301 cricetus longica
seq_name: SwissProt_39:ECIC_BOVIN
seq_documentation_block:
ID ECIC_BOVIN STANDARD; PRT; 903 AA.
AC P54281;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN (CALCIUM-ACTIVATED CHLORIDE
DE CHANNEL).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Trachea;
RX MEDLINE=96125078; PubMed=8537359;
RA Cunningham S.A., Awayda M.S., Bubien J.K., Ismailov I.I.,
RA Arrate M.P., Berdiev B.K., Benos D.J., Fuller C.M.;
RT "Cloning of an epithelial chloride channel from bovine trachea.";
RL J. Biol. Chem. 270:31016-31026(1995).
CC -!- FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE
CC SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME;
CC MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND
CC TRANSEPITHELIAL TRANSPORT.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- TISSUE SPECIFICITY: TRACHEA.
CC -!- PTM: PHOSPHORYLATED BY CAM-KINASE II.
CC -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U36445; AAC48511.1; -
CC InterPro; IPR002035; VWFA.
CC SMART; SM00327; VWFA; 1.
CC PROSITE; PS50234; VWFA; 1.
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW Calcium channel; Phosphorylation; Glycoprotein.
FT TRANSMEM 7 27
FT TRANSMEM 331 351
FT TRANSMEM 617 637
FT TRANSMEM 883 903
FT DOMAIN 308 476
FT CARBOHYD 75 75
FT CARBOHYD 278 278
FT CARBOHYD 360 360
FT CARBOHYD 372 372
FT CARBOHYD 504 504
FT CARBOHYD 515 515
FT CARBOHYD 688 688
FT CARBOHYD 811 811
FT CARBOHYD 816 816
FT CARBOHYD 842 842
FT CARBOHYD 857 857
SQ SEQUENCE 903 AA; 100305 MW; 9742AB1590908AEC CRC64;
alignment_scores:
Quality: 83.50 Length: 40

Ratio: 2.783 Gaps: 1
Percent similarity: 75.000 Percent identity: 45.000

alignment_block:

US-09-049-696-15 x ECLC_BOVIN ..

Align seg 1/1 to: ECLC_BOVIN from: 1 to: 903

1 ACAGATCTTTTCATCTGCTATTCAGGCTGTGATAAGTCGATCTGAAATC 50

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

844 ThrAspPheTyrIleAlaValGlnAlaIleAsnGluAlaAsnLeuThrSe 860

51 AGAAATATCCAAATTCGACGAGTATCTTTGTTTATTCCTCCACAGACTC 100

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

860 rGluValSerAsnIleAlaGlnAlaIleTyrPheIlePro.....M 874

101 CGCCAGACGACCTAGTCCT 120

|||||

874 etProGluAspSerValPro 880

seq_name: SwissProt_39:SUTS_RAT

seq_documentation_block:

ID SUTS_RAT STANDARD: PRT: 1840 AA.

AC P23739;

DT 01-NOV-1991 (Rel. 20, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE SUCRASE-ISOMALTASE, INTESTINAL [CONTAINS: SUCRASE (EC 3.2.1.48);

DE ISOMALTASE (EC 3.2.1.10)]

GN SI.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Intestine;

RX MEDLINE=95121929; PubMed=7821806;

RA Chandrasena G., Osterholm D.E., Sunitha I., Henning S.J.;

RT "Cloning and sequencing of a full-length rat sucrase-isomaltase-

RT encoding cDNA.";

RL Gene 150:355-360(1994).

RN [2]

RP SEQUENCE OF 86-361 FROM N.A.

RC STRAIN=FISCHER 344; TISSUE=Intestine;

RX MEDLINE=91097578; PubMed=2268340;

RA Traber P.G.;

RT "Regulation of sucrase-isomaltase gene expression along the crypt-

RT villus axis of rat small intestine.";

RL Biochem. Biophys. Res. Commun. 173:765-773(1990).

RN [3]

RP SEQUENCE OF 732-1372 FROM N.A.

RC STRAIN=SPRAGUE-DAWLEY; TISSUE=DuoDenum;

RX MEDLINE=90381315; PubMed=2400788;

RA Brody J.-P., Hugot J.-P., Perret C., Porteu A.;

RT "Molecular cloning and characterization of a rat intestinal sucrase-

RT isomaltase cDNA. Regulation of sucrase-isomaltase gene expression by

RT sucrose feeding.";

RL Biochim. Biophys. Acta 1087:61-67(1990).

RN [4]

RP SEQUENCE OF N-TERMINUS OF ISOMALTASE AND SUCRASE.

RX MEDLINE=82167342; PubMed=6802834;

RA Hauri H.-P., Wacker H., Rickli E.E., Bigler-Meyer B., Quaroni A.,

RA Semenza G.;

RT "Biosynthesis of sucrase-isomaltase. Purification and NH2-terminal

RT amino acid sequence of the rat sucrase-isomaltase precursor

RT (pro-sucrase-isomaltase) from fetal intestinal transplants.";

RL J. Biol. Chem. 257:4522-4528(1982).

CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE FINAL STAGE OF

CC CARBOHYDRATE DIGESTION.

CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF SUCROSE AND MALTOSE BY AN

CC

CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,6-ALPHA-D-GLUCOSIDIC LINKAGES
CC IN ISOMALTOSE AND DEXTRINS PRODUCED FROM STARCH AND GLYCOGEN BY
CC ALPHA-AMYLASE.
CC -!- SUBUNIT: THE RESULTING SUCRASE AND ISOMALTASE SUBUNITS STAY
CC ASSOCIATED WITH ONE ANOTHER IN A COMPLEX BY NON-COVALENT LINKAGES.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. BRUSH BORDER.
CC -!- PTM: THE PRECURSOR IS PROTEOLYTICALLY CLEAVED WHEN EXPOSED TO
CC PANCREATIC PROTEASES IN THE INTESTINAL LUMEN.
CC -!- PTM: SULFATED (BY SIMILARITY).
CC -!- MISCELLANEOUS: THERE IS A HIGH DEGREE OF HOMOLOGY BETWEEN THE
CC ISOMALTASE AND SUCRASE PORTIONS (41 % OF AMINO ACID IDENTITY)
CC INDICATING THAT THIS PROTEIN IS EVOLVED BY PARTIAL GENE
CC DUPLICATION.
CC -!- SIMILARITY: CONTAINS 1 P-TYPE (TREFOL) DOMAIN.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L25926; AAA65097.1; -;
CC EMBL; M62889; AAA42144.1; -;
CC EMBL; X15546; CAA33552.1; -;
CC PIR; S11386; S11386.
CC InterPro: IPR000322; Glyco_hydro_31.
CC InterPro: IPR000519; P_trefol.
CC Pfam; PF01055; Glyco_hydro_31; 2.
CC Pfam; PF00088; trefol; 1.
CC SMART; SM00018; P; 2.
CC PROSITE; PS00025; P_TREFOL; 1.
CC PROSITE; PS00129; GLYCOSYL_HYDROL_F31_2; FALSE_NEG.
CC PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2; FALSE_NEG.
CC KW Multifunctional enzyme; Transmembrane; Glycoprotein; Hydrolase;
CC Glycosidase; Repeat; Signal-anchor; Sulfation.
CC INIT_MET 0 0
CC CHAIN 1 1840 SUCRASE-ISOMALTASE, INTESTINAL.
FT CHAIN 1 1012 ISOMALTASE.
FT CHAIN 1013 1840 SUCRASE.
FT DOMAIN 1 11 CYTOPLASMIC.
FT DOMAIN 12 31 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 32 1840 LUMENAL.
FT DOMAIN 42 69 SER/THR-RICH.
FT DOMAIN 71 118 P-TYPE.
FT DOMAIN 119 1012 ISOMALTASE.
FT DOMAIN 1013 1840 SUCRASE.
FT ACT_SITE 513 513 BY SIMILARITY.
FT ACT_SITE 1398 1398 BY SIMILARITY.
FT DISULFID 72 103 BY SIMILARITY.
FT DISULFID 86 102 BY SIMILARITY.
FT DISULFID 97 115 BY SIMILARITY.
FT MOD_RES 400 400 SULFATION (POTENTIAL).
FT MOD_RES 409 409 SULFATION (POTENTIAL).
FT MOD_RES 1386 1386 SULFATION (POTENTIAL).
FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 757 757 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 866 866 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 909 909 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1239 1239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1307 1307 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1344 1344 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1358 1358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1372 1372 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1484 1484 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1512 1512 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1574 1574 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1761 1761 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1828 1828 N-LINKED (GLCNAC. . .) (POTENTIAL).

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
TISSUE=Brain;
MEDLINE=95048367; PubMed=7959765;
RA Brookman K.W., Tebbes R.S., Allen S.A., Tucker J.D., Swiger R.R.,
Kamerlain J.E., Carrano A.V., Thompson L.H.;
"Isolation and characterization of mouse Xrcc-1, a DNA repair gene
affecting ligation.";
Genomics 22:180-188(1994).
CC -1- FUNCTION: CORRECTS DEFECTIVE DNA SPAND-BREAK REPAIR AND SISTER
CC -2- CHROMATID EXCHANGE FOLLOWING TREATMENT WITH IONIZING RADIATION
CC -3- AND ALKYLATING AGENTS.
CC -4- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -5- SIMILARITY: SOME, TO S.POMBE RAD4/CUT5.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U02887; AAA93115.1; -
DR MGD; MGI:99137; XRCCL.
DR InterPro; IPR001357; BRCT.
DR InterPro; IPR002706; XRCCL_N.
DR Pfam; PF00533; BRCT; 2.
DR Pfam; PF01834; XRCCL_N; 1.
DR ProDom; PD023136; XRCCL_N; 1.
DR SMART; SM00292; BRCT; 2.
DR PROSITE; PSS0172; BRCT; 2.
DR DNA repair; Nuclear protein.
DR KWA
SQ SEQUENCE 631 AA; 69003 MW;
7F1868BFE2BA3C68 CRC64;

```
alignment_scores:
  Quality: 62.00      Length: 18
  Ratio: 4.133       Gaps: 0
  Percent Similarity: 83.333  Percent Identity: 55.556
```

```
alignment_block:
    "JS-09-049-696-15 x XRCL_MOUSE ..
    :Align{seq l/v} to: XRCL_MOUSE from: 1 to: 631
```

[illegible]

```

seq_name: SwissProt_39:BCHD_HELMO
seq_documentation_block:
ID      BCHD_HELMO      STANDARD;      PRT;      666 AA.
AC      Q9ZGE6;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      MAGNESIUM-CHELATASE 67 KDA SUBUNIT (MG-PROTOPORPHYRIN IX CHELATASE).
DE      (MG-CHELATASE SUBUNIT D).
GN      BCHD.
OS      Helicobacillus mobilis.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group;
GC      Helicobacterium group; Helicobacillus.
CC      NCBI_Taxid=28064;
[1]
RN
RP      SEQUENCE FROM N.A.

```


RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers J.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottiler P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleeb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -!- FUNCTION: TRANSFERS THE GLYCOSYL RESIDUE FROM UDPG TO THE
 CC NONREDUCING END OF ALPHA-1,4-GLUCAN.
 CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + (1,4-ALPHA-D-GLUCOSYL)[N] = UDP
 CC + (1,4-ALPHA-D-GLUCOSYL)[N+1].
 CC -!- PATHWAY: GLYCODEN BIOSYNTHESIS.
 CC -!- SIMILARITY: BELONGS TO THE MAMMALIAN/FUNGAL GLYCOSYL SYNTHASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE003707; AAF55132.1;
 DR FlyBase: FBgn0038293; CG6904.
 KW Hypothetical protein; Glycogen biosynthesis; Transferase;
 KW Glycosyltransferase.
 FT BINDING 41 41 UDP-GLUCOSE (BY SIMILARITY).
 SQ SEQUENCE 689 AA; 79233 MW; F29348548680B2A CRC64;

alignment_scores:
 Quality: 60.50 Length: 51
 Ratio: 1.779 Gaps: 2
 Percent Similarity: 66.667 Percent Identity: 27.451

alignment_block:

US-09-049-696-15 x GYS_DROME

Align seg 1/1 to: GYS_DROME from: 1 to: 689

16 GCTATTACGGCT.....GTTGTAAGGTCGATCTGAATC 50

623 AlaLeuGlnAlaValTyrProAspTyrValAspGluLeuSerLeuTyrG1 639
 51 AGAAATATCCAACTGCACGAGTATCTTTGTTTATTCCTCCACGACACTC 100
 639 YSerlyAsnAsnLeu.....IlePheSerArgProHisSerG 652
 101 CCCAGAGACACTAGTCTCTGATGAACGCTGCTCTTGTCTCTATATT 150
 652 luProSerProThrSerSerArgHisThrThrProAlaProSerVal 668
 151 CAT 153
 669 His 669
 Seq_name: SwissProt_39:FIMC_SALTY
 seq_documentation_block:
 ID FIMC_SALTY STANDARD; PRT; 230 AA.
 AC P74923;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CHAPERONE PROTEIN FIMC PRECURSOR.
 GN FIMC.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 CC NCBI_TaxID=602;
 RN [?].
 RP SEQUENCE FROM N.A.
 RA Swenson D.L., Clegg S.;
 RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: REQUIRED FOR THE BIOGENESIS OF TYPE 1 FIMBRIAE.
 CC BINDS AND INTERACT WITH FIMH.
 CC -!- SUBCELLULAR LOCATION: PERIPLASMIC (POTENTIAL).
 CC -!- SIMILARITY: BELONGS TO THE PERIPLASMIC PILUS CHAPERONE FAMILY.
 CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L19338; AAA75418.1;
 DR StGene: SG10504; fmc.
 DR InterPro: IPR001829; Pili_chaperone.
 DR Pfam: PF00345; pili_assembly; 1.
 DR PRINTS: PR00969; CHAPERONPILI.
 DR PRODOM: PD001447; Pili_chaperone; 1.
 DR PROSITE: PS00635; PILI_CHAPERONE; 1.
 KW Chaperone; Fimbria; Periplasmic; Signal; Immunoglobulin domain.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 230 CHAPERONE PROTEIN FIMC.
 SQ SEQUENCE 230 AA; 24867 MW; 9CC193826E92A712 CRC64;

alignment_scores:
 Quality: 60.00 Length: 42
 Ratio: 2.143 Gaps: 1
 Percent Similarity: 66.667 Percent Identity: 30.952

alignment_block:

US-09-049-696-15 x FIMC_SALTY

Align seg 1/1 to: FIMC_SALTY from: 1 to: 230

16 GCTATTACGGCTGTTGTAAGTCGATCTGAATCA.....GA 53

113 AlaIleProSerValAspLysSerHisIleGluGlyAsnValLeuG1 129

54 AATATCCAACTGACGATCTTTGTTTATCTCTCCACAGACTCCG 103
nleuAlaIleLeuSerArgIleuLeuValArgProAlaAsnLeuP 146
129 CAGAGACACTAGCTGTGATGAACG 129
146 roGlnThrProGluAspAlaProThr 154

seq_name: SwissProt_39:CDNC_MOUSE

seq_documentation_block:
ID CDNC_MOUSE STANDARD; PRT; 348 AA.
AC P49319;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE CYCLIN-DEPENDENT KINASE INHIBITOR IC (CYCLIN-DEPENDENT KINASE
DE INHIBITOR P57) (P57KIP2).
GN CDKN1C OR KIP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=95247027; PubMed=7729683;
RA Lee M.-H., Reynolds J. I., Massague J.;
RT "Cloning of p57KIP2, a cyclin-dependent kinase inhibitor with unique
RT domain structure and tissue distribution.";
RL Genes Dev. 9:639-649(1995).
RN [2]
RP SEQUENCE OF 14-348 FROM N.A.
RX MEDLINE=95247028; PubMed=7729684;
RA Matsuo S., Edwards M.C., Bai C., Parker S., Zhang P., Baldini A.,
RA Harper J.W., Elledge S.J.;
RT "p57KIP2, a structurally distinct member of the p21CIP1 Cdk inhibitor
RT family, is a candidate tumor suppressor gene.";
RL Genes Dev. 9:650-662(1995).
CC -1- FUNCTION: POTENT TIGHT-BINDING INHIBITOR OF SEVERAL G1 CYCLIN/CDK
CC COMPLEXES (CYCLIN E-CDK2, CYCLIN D2-CDK4, AND CYCLIN A-CDK2) AND,
CC TO LESSER EXTENT, OF THE MITOTIC CYCLIN B-CDK2. NEGATIVE REGULATOR
CC OF CELL PROLIFERATION. MAY PLAY A ROLE IN MAINTENANCE OF THE
CC NONPROLIFERATIVE STATE THROUGHOUT LIFE.
CC SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: IT IS EXPRESSED IN THE HEART, BRAIN, LUNG,
CC SKELETAL MUSCLE, KIDNEY, PANCREAS AND TESTIS. HIGH LEVELS ARE SEEN
CC IN THE PLACENTA WHILE LOW LEVELS ARE SEEN IN THE LIVER.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U20553; AAC52186.1;
EMBL; U22399; AAA85096.1;
DR InterPro: IPR003175; CDK1.
DR Pfam; PF02234; CDI; 1.
KW Cell cycle; Alternative splicing.
FT DOMAIN 108 189 PRO-RICH.
FT DOMAIN 178 284 GLU/ASP-RICH.
FT DOMAIN 309 312 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT VARSPLIC 1 13 MISSING (IN ISOFORM KIP2B/P57B).
FT CONFLICT 150 151 DA -> EP (IN REF. 2).
SQ SEQUENCE 348 AA; 37331 MW; 108A8538D77016D9 CRC64;

alignment_scores:

Quality: 60.00 Length: 45
Ratio: 2.500 Gaps: 1
Percent Similarity: 53.333 Percent Identity: 35.556
alignment_block:
US-09-049-696-15 x CDNC_MOUSE
Align seg 1/1 to: CDNC_MOUSE from: 1 to: 348
22 CAGGCTGTTGATAGTTCGATCTGAAATCAGAAATATCCAACTGTCACG 71
127 GluAlaProAspGlyLeuGluAlaProGluGlnProProSerAlaPr 143
72 AGTATCTTTGTTTATCTCTCCACAGACTCCGCCACAGACACTAGTCTCT. 120
143 oAlaSerAlaValAlaAspAlaThrProProAlaThrProAlaProA 160
121GATGAACGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT. 150
160 laSerAspLeuThrSerAspProIleProGluVal 171

seq_name: SwissProt_39:APG_ARATH

seq_documentation_block:

ID APG_ARATH STANDARD; PRT; 534 AA.
AC P49602;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE ANTER-SPECIFIC PROLINE-RICH PROTEIN APG PRECURSOR.
GN APG.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94004980; PubMed=8401599;
RA Roberts M.R., Foster G.D., Blundell R.P., Robinson S.W., Kumar A.,
RA Draper J., Scott R.;
RT "Gametophytic and sporophytic expression of an anther-specific
RT Arabidopsis thaliana gene.";
RL Plant J. 3:111-120(1993).
CC -1- TISSUE SPECIFICITY: FOUND IN SPOROPHYTIC AND GAMETOPHYTIC CELL
CC TYPES IN THE ANTER, ONLY IN MALE FERTILE PLANTS.
CC DEVELOPMENTAL STAGE: EXPRESSED IN MALE GAMETOGENESIS. DURING
CC MICROSPORE DEVELOPMENT. HIGHER EXPRESSION IS FOUND DURING
CC MICROSPORE MITOSIS WITH A DRAMATIC DECLINE DURING POLLEN
CC MATURATION.
CC -1- SIMILARITY: BELONGS TO THE "GDSL" FAMILY OF LIPOLYTIC ENZYMES.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X60377; CAA42925.1;
DR FIR; S21961; S21961.
DR InterPro: IPR001087; Lipase_GDSL.
DR Pfam; PF00657; Lipase_GDSL; 1.
DR PROSITE; PS01098; LIPASE_GDSL_SER; 1.
KW Signal.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 534 ANTER-SPECIFIC PROLINE-RICH PROTEIN APG.
FT ACT_SITE 211 211 BY SIMILARITY.
FT ACT_SITE 511 511 POTENTIAL.
SQ SEQUENCE 534 AA; 57967 MW; 744CAD3B08CC482E CRC64;

alignment_scores: Quality: 60.00 Length: 20
Ratio: 4.615 Gaps: 0
Percent Similarity: 65.000 Percent Identity: 50.000
alignment_block:
US-09-049-696-15 x APG_ARATH ..
Align seg 1/1 to: APG_ARATH from: 1 to: 534
85 ATTCTCCACAGACTCCGACAGACACTAGTCTGATGAAGCTGTGC 134
:::|||||: ||||| ||||| |||||
169 ValProHisGlyProProlyProAlaProAlaProThrProAl 185
135 TCCTGTGCT 144
||||| |||
185 aProSerPro 188

seq_name: SwissProt_39:EIA_ADE07

seq_documentation_block:
ID EIA_ADE07 STANDARD; PRT; 261 AA.
AC P03256;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE EARLY EIA 28 KDA PROTEIN [CONTAINS: EARLY EIA 24 KDA PROTEIN].
OS Human adenovirus type 7.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=10519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMEN;
RX MEDLINE=81237792; PubMed=6985480;
RA Dijkema R., Dekker B.M.M., van Ormondt H., de Waard A., Maat J.,
RA Boyer H.W.;
RT "Gene organization of the transforming region of weakly oncogenic
RT adenovirus type 7: the E1a region.";
RL Gene 12:287-299(1980).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GRIDER;
RA Yoshida K., Fujinaga K.;
RT "The nucleotide sequence of the transforming HindIII-I.J fragment of
RT adenovirus type 7 DNA.";
RL Tumor Res. 19:39-47(1984).
CC -1- FUNCTION: TRANS-ACTIVATES EARLY VIRAL PROMOTERS AND SOME CELLULAR
CC PROMOTERS.
CC -1- ALTERNATIVE PRODUCTS: THESE PROTEINS AND ONE OTHER, THE EARLY 6.3
CC KDA PROTEIN, ARE PRODUCTS OF THREE DIFFERENT MRNAS SPLICED FROM
CC THE SAME PRIMARY TRANSCRIPT FROM THE E1A REGION OF THE GENOME.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03000; CAA26758.1; -
DR EMBL; X03000; CAA26759.1; -
DR EMBL; M38648; AAA2454.1; -
DR EMBL; M38648; AAA2455.1; -
DR PIR; A03826; WMAD87.
DR InterPro; IPR003853; Adeno.E1A.
DR Pfam; PF02703; Adeno.E1a; 1.
KW Transcription regulation; Activator; Early protein;
KW Alternative splicing; Zinc-finger; DNA-binding.
FT ZN_FING 163 183
FT VARSPIC 164 194
FT VARIANT 62 62
FT VARIANT 202 202

SQ SEQUENCE 261 AA; 28385 MW; 770E327BD524A06C CRC64;
alignment_scores: Quality: 59.50 Length: 30
Ratio: 2.705 Gaps: 3
Percent Similarity: 73.333 Percent Identity: 50.000
alignment_block:
US-09-049-696-15 x EIA_ADE07 ..
Align seg 1/1 to: EIA_ADE07 from: 1 to: 261
63 CATTGCACGAGTATCTTTGTTTATTCCTCCACAGACTCCGACAGACAC 112
:::|||||: ||||| ||||| |||||
188 HisCysHis.....PheIleTySer.ProValSerAspGluSerP 202
113 CTAGTCTGATGAACGTCCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 150
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
202 roserProAspSerThrThrSerPro...ProGluile 213
seq_name: SwissProt_39:DIAL_HUMAN

seq_documentation_block:
ID DIAL_HUMAN STANDARD; PRT; 1248 AA.
AC O60610; Q90C76;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE DIAPHANOUS PROTEIN HOMOLOG 1 (DIAPHANOUS-RELATED FORMIN 1) (DRF1).
GN DIAPHI OR DIAP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
KN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98028756; PubMed=9360932;
RA Lynch E.D., Lee M.K., Morrow J.E., Welch P.L., Leon P.E., King M.-C.;
RT "Nonsyndromic deafness DFNA1 associated with mutation of a human
RT homolog of the Drosophila gene diaphanous.";
RL Science 278:1315-1318(1997).
RN [2]
RP SEQUENCE OF 218-817 FROM N.A.
RC TISSUE=Ovarian carcinoma;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RN Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 727-765 AND 1121-1145.
RC TISSUE=Platelet;
RX MEDLINE=95255215; PubMed=7737110;
RA Reinhard M., Giehl K., Abel K., Haffner C., Jarchau T., Hoppe V.,
RA Jockusch B.M., Walter U.;
RT "The proline-rich focal adhesion and microfilament protein VASP is a
RT ligand for profilins.";
RL EMBL J. 14:1583-1589(1995).
CC -1- FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN
CC A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE
CC IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,
CC STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE
CC SERUM RESPONSE FACTOR. DFR PROTEINS COUPLE RHO AND SRC TYROSINE
CC KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS (BY
CC SIMILARITY). IN HEARING IT MAY PLAY A ROLE IN THE REGULATION OF
CC ACTIN POLYMERIZATION IN HAIR CELLS.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART, PLACENTA, LUNG,
CC KIDNEY, PANCREAS, LIVER, SKELETAL MUSCLE, AND COCHLEA.

CC -|- DOMAIN: DRFS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
CC RHO-GTP ACTIVATES THE DRFS BY DISRUPTING THE GBD-DAD INTERACTION
CC (BY SIMILARITY).
CC -|- DISEASE: DEFECTS IN DIAPH1 ARE A CAUSE OF AUTOSOMAL DOMINANT
CC NONSyndromic SENSORINEURAL DEAFNESS 1 (DFNA1).
CC -|- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).
CC -|- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FH1) DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 3 (FH3) DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 DRF AUTOREGULATORY DOMAIN (DAD).
CC -|- SIMILARITY: CONTAINS 1 DRF AUTOREGULATORY DOMAIN (DAD).
CC -|- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. DIAPHANOUS
CC SUBFAMILY.
CC -|- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THAT IT SEEMS
CC TO INCLUDE INTRONIC SEQUENCE.
CC -|- DATABASE: NAME=hereditary hearing loss homepage;
CC NOTE=Gene page;
CC WWW="http://www.uia.ac.be/dnalab/hhh/hhhgenes.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF051782; AAC05373.1; .
DR EMBL; AK023345; BAB14533.1; ALT_SEQ.
DR MIM; 602121; .
DR MIM; 124900; .
DR InterPro; IPR003104; FH2.
DR InterPro; IPR002965; P_Rich_extensn.
DR Pfam; PF02181; FH2; 1.
DR SMART; SM00498; FH2; 1.
KW Coiled coil; Repeat; Deafness.
FT DOMAIN 63 260
FT DOMAIN 157 457
FT DOMAIN 460 563
FT DOMAIN 588 743
FT DOMAIN 748 1190
FT DOMAIN 1015 1172
FT DOMAIN 1173 1187
FT DOMAIN 1189 1192
FT DOMAIN 1189 1192 ARG/LYS-RICH (BASIC).
FT CONFLICT 804 804 T -> TSKA (IN REF. 2).
FT CONFLICT 1132 1133 RK -> AE (IN REF. 3).
SQ SEQUENCE 1248 AA; 138978 MW; EDIF5147CFF9A886 CRC64;

alignment_scores: Quality: 59.50 Length: 54
Ratio: 2.125 Gaps: 1
Percent Similarity: 51.852 Percent Identity: 20.370
alignment_block:
US-09-049-696-15 x DIAL_HUMAN ..
Align seg 1/1 to: DIAL_HUMAN from: 1 to: 1248
40 GATCTGAATCAGAAATATCCACATTCGACGATATCTTTGTTATTCC 89
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
548 AspalalysLysGluMetAlaSerLeuSerAlaAlaAlaIleThrValPr 564
90 TCCACAGACTCCG..... 102
|||||:|||||
564 oProSerValProSerArgAlaProValProProAlaProLeuProG 581
103CCAGAGACACTAGTCTCTGATGAACAGCTCT 132
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
581 lYAspSerGlyThrIleIleProProProAlaProGlyAspSerThr 597
133 GCTCCTTGCTCT 144
:|||||
598 ThrProProPro 601

seq_name: SwissProt_39:PGCA_CHICK
seq_documentation_block:
ID PGCA_CHICK STANDARD; PRT: 2109 AA.
AC P07898; Q90991; Q90820; Q91047; Q90810;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE AGGREGAN CORE PROTEIN PRECURSOR (CARTILAGE-SPECIFIC PROTEOGLYCAN CORE
DE PROTEIN) (CSPC).
GN AGG1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHITE LEGHORN; TISSUE=Embryo;
RX MEDLINE=94043149; PubMed=8226878;
PA Li H., Schwartz N.B., Vertel B.M.;
FT "cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core
FT protein and identification of a stop codon in the aggrecan gene
FT associated with the chondrodystrophy, nanomelia.";
FT J. Biol. Chem. 268:23504-23511(1993).
RN [2]
RP SEQUENCE OF 1042-1559 FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=90307744; PubMed=1694853;
PA Krieger R.C. Jr., Fields T.A., Mensch J.R. Jr., Schwartz N.B.;
FT "Chick cartilage chondroitin sulfate proteoglycan core protein. II.
FT Nucleotide sequence of cDNA clone and localization of the S103L
FT epitope.";
FT J. Biol. Chem. 265:12088-12097(1990).
RN [3]
RP SEQUENCE OF 1-1855 AND 1893-2109 FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=93111968; PubMed=1339285;
PA Chandrasekaran L., Tanzer M.L.;
FT "Molecular cloning of chicken aggrecan. Structural analyses.";
RN [4]
RP ERRATUM.
RX MEDLINE=94107258; PubMed=8280087;
PA Chandrasekaran L., Tanzer M.L.;
RL Biochem. J. 296:885-887(1993).
RN [5]
RP SEQUENCE OF 1492-1610 FROM N.A.
RC STRAIN=WHITE LEGHORN; TISSUE=Chondrocytes;
RX MEDLINE=95128519; PubMed=7827752;
PA Primorac D., Stover M.L., Clark S.H., Rowe D.W.;
FT "Molecular basis of nanomelia, a heritable chondrodystrophy of
FT chicken.";
RL Matrix Biol. 14:297-305(1994).
RN [6]
RP SEQUENCE OF 1894-2109 FROM N.A.
RX MEDLINE=89008500; PubMed=3170613;
PA Tanaka T., Har-El R., Tanzer M.L.;
FT "Partial structure of the gene for chicken cartilage proteoglycan
FT core protein.";
RL J. Biol. Chem. 263:15831-15835(1988).
RN [7]
RP SEQUENCE OF 1693-1855 AND 1893-2109 FROM N.A.
RX MEDLINE=86259736; PubMed=3460082;
PA Sai S., Tanaka T., Koshier R.A., Tanzer M.L.;
FT "Cloning and sequence analysis of a partial cDNA for chicken
FT cartilage proteoglycan core protein.";
RL proc. Natl. Acad. Sci. U.S.A. 83:5081-5085(1986).
CC -|- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A

REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY
SIMILARITY).

-1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
THE A, B, C MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
AND G3.

-1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.

-1- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF NANOMELIA, A LETHAL
CONNECTIVE TISSUE DISORDER AFFECTING CARTILAGE DEVELOPMENT
(CHONDRODYSPLASIA) CHARACTERIZED BY SHORTENED AND MALFORMED LIMBS.
AGGREGAN IS TRUNCATED AT ITS C-TERMINAL IN THE CS-2 BINDING DOMAIN
AND IS NOT ANYMORE SECRETED FROM THE CHONDROCYTES.

-1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

-1- SIMILARITY: CONTAINS 4 LINK DOMAINS.

-1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.

-1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

-1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.

-1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L21913; AAB19128.1; -;
EMBL; M38187; AAA48731.1; -;
EMBL; M88101; -; NOT_ANNOTATED_CDS.
EMBL; S74657; AAC60751.1; -;
EMBL; S74656; AAC60751.1; JOINED.
EMBL; J04028; AAA48719.1; -;
EMBL; M33993; AAA48720.1; -;
PIR; A25442; A25442.
HSP; P00740; 11XA.
InterPro: IPR000152; Asx_hydroxyl.
InterPro: IPR000561; EGF-like.
InterPro: IPR000742; EGF_2.
InterPro: IPR001881; EGF_Ca.
InterPro: IPR003599; Ig.
InterPro: IPR000538; Link.
InterPro: IPR003324; SGXSG.
InterPro: IPR00436; Sushi_SCR_CCP.
InterPro: IPR001304; lectin_c.
Pfam; PF00008; EGF; 1.
Pfam; PF00059; lectin_c; 1.
Pfam; PF02339; SGXSG; 56.
Pfam; PF00084; sushi; 1.
Pfam; PF00193; xlink; 4.
ProDom; PD000918; Link; 4.
SMART; SM00032; CCP; 1.
SMART; SM00034; CLECT; 1.
SMART; SM00179; EGF_CA; 1.
SMART; SM00409; IG; 1.
SMART; SM00445; LINK; 4.
PROSITE; PS00010; ASX_HYDROXYL; 1.
PROSITE; PS00022; EGF_1; 1.
PROSITE; PS01187; EGF_CA; 1.
PROSITE; PS01241; LINK; 4.
PROSITE; PS00615; C-TYPE LECTIN_1; 1.
PROSITE; PS50041; C-TYPE LECTIN_2; 1.
Glycoprotein; Cartilage; Proteoglycan; Lectin; Sushi; Sushii;
EGF-like domain; Alternative splicing; Repeat; Immunoglobulin domain.
SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 2109 AGGREGAN CORE PROTEIN.
FT DOMAIN 44 136 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 166 243 LINK 1.

DOMAIN	264	346	LINK 2.
FT DOMAIN	537	614	LINK 3.
FT DOMAIN	635	716	LINK 4.
FT DOMAIN	1363	1742	19 X 20 AA TANDEM-REPEAT.
FT DOMAIN	1855	1892	EGF-LIKE.
FT DOMAIN	1901	2019	C-TYPE LECTIN.
FT DOMAIN	2023	2081	SUSHI.
FT DOMAIN	48	137	G1-A.
FT DOMAIN	148	243	G1-B.
FT DOMAIN	249	346	G2-B.
FT DOMAIN	519	613	G2-B.
FT DOMAIN	620	715	G2-B.
FT DOMAIN	718	803	KS.
FT DOMAIN	805	1264	CS-1.
FT DOMAIN	1265	1742	CS-2.
FT DOMAIN	1893	2109	G3.
FT DISULFID	51	129	BY SIMILARITY.
FT DISULFID	171	242	BY SIMILARITY.
FT DISULFID	195	216	BY SIMILARITY.
FT DISULFID	269	345	BY SIMILARITY.
FT DISULFID	293	314	BY SIMILARITY.
FT DISULFID	542	613	BY SIMILARITY.
FT DISULFID	566	587	BY SIMILARITY.
FT DISULFID	640	715	BY SIMILARITY.
FT DISULFID	664	685	BY SIMILARITY.
FT DISULFID	1859	1870	BY SIMILARITY.
FT DISULFID	1864	1879	BY SIMILARITY.
FT DISULFID	1881	1890	BY SIMILARITY.
FT DISULFID	1897	1908	BY SIMILARITY.
FT DISULFID	1925	2017	BY SIMILARITY.
FT DISULFID	1993	2009	BY SIMILARITY.
FT DISULFID	2024	2067	BY SIMILARITY.
FT DISULFID	2053	2080	BY SIMILARITY.
FT CARBOHYD	76	76	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	122	122	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	330	330	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	388	388	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	439	439	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	644	644	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	700	700	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	765	765	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	801	801	N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPIC	1856	1892	MISSING (IN ISOFORM 2).
FT CONFLICT	362	362	E -> D (IN REF. 3).
FT CONFLICT	601	601	G -> D (IN REF. 3).
FT CONFLICT	1000	1000	P -> R (IN REF. 3).
FT CONFLICT	1029	1029	A -> P (IN REF. 3).
FT CONFLICT	1042	1043	VT -> PA (IN REF. 2).
FT CONFLICT	1251	1251	E -> D (IN REF. 2 AND 3).
FT CONFLICT	1587	1587	I -> T (IN REF. 5).
FT CONFLICT	1590	1590	I -> V (IN REF. 5).
FT CONFLICT	1594	1594	T -> S (IN REF. 5).
FT CONFLICT	1602	1610	1ETSTVREI -> VLRCRSVLR (IN REF. 5).
FT CONFLICT	1603	1603	E -> A (IN REF. 3).
FT CONFLICT	1672	1672	S -> G (IN REF. 3).
FT CONFLICT	1796	1796	E -> G (IN REF. 3 AND 7).
FT CONFLICT	1988	1988	F -> S (IN REF. 6).
SO SEQUENCE	2109 AA; 223492 MW; 7F824FD5B3A2ABDA CRC64;		

alignment_scores:
Quality: 59.50 Length: 50
Ratio: 2.052 Gaps: 1
Percent Similarity: 58.000 Percent Identity: 36.000

alignment_block:

US-09-049-696-15 x PGCA_CHICK

Align:seg 1/1 to: PGCA_CHICK from: 1 to: 2109

4 GATCTTTTCATTGCTATTTCAGCTGTTTCATAGGTCGATCTCAATCAGA 53
||||| :||| ||| ||| :||| :|||
176i AspAlaValIleSerThrSerAlaProasp...ValGluLeuAlaGlnG1 1796

54 AATATCAACATTGCACGAGTATCTTTGTTTATCTCCACAGACTCCGC 103
1796 uProArgAsnThrGluThrGlnLeuGluThrGluProSerThrProA 1813
104 CAGAGACACTAGTCTGTGATGAACGTGCTCCTGTCTATATTCAT 153
1813 laAlaSerGlyGlnGluThrGluThrAlaAlaValLeuAspAsnProHis 1829
seq_name: SwissProt_39:NU5M_CHICK

seq_documentation_block:
ID NUSM_CHICK STANDARD; PRT; 605 AA.
AC P18940;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).
GN MTND5 OR ND5 OR NADH5.
OS Gallus gallus (Chicken).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90230301; PubMed=2329578;
RA Desjardins P., Morais R.;
RT "Sequence and gene organization of the chicken mitochondrial genome.
RT A novel gene order in higher vertebrates.";
RL J. Mol. Biol. 212:599-634(1990).
CC -/- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X52392; CAA36635.1; .
DR PIR: S10197.
DR InterPro: IPR003916; NADHub_oxdrdctse5.
DR InterPro: IPR001750; Oxidored_q1.
DR InterPro: IPR001516; Oxidored_q1_N.
DR Pfam: PF00361; oxidored_q1; 1.
DR Pfam: PF00662; oxidored_q1_N; 1.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 605 AA; 66373 MW; BC150F8693131E0F CRC64;

alignment_scores:
Quality: 58.50 Length: 41
Ratio: 2.089 Gaps: 1
Percent Similarity: 68.293 Percent Identity: 36.585

alignment_block:

US-09-049-696-15 x NU5M_CHICK ..
Align seg 1/1 to: NU5M_CHICK from: 1 to: 605

16 GCTATTGAGGCTGTTGATAGGTCGATCGAATCAGAAATATCCACAT 65
450 AlAlaLeuProIleMetArgLeuAlaLeuGlySerIleMetAlaGlyLe 466
56 TGCACGAGTATCTTTGTTTATCTCCACAGACTCCGCCAGAGACA...C 112
466 uLeuIleSerSerLeuIleLeuProLysThrProProMetThrMetP 483
113 CTACTCTGTGATGAACGTCGCT 135
483 roThrIleThrLysThrAlaAla 490

seq_name: SwissProt_39:YAR2_SCHPO

seq_documentation_block:
ID YAR2_SCHPO STANDARD; PRT; 1273 AA.
AC Q10135;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HYPOTHETICAL 142.5 KDA PROTEIN C23E2.02 IN CHROMOSOME 1.
GN SPAC23E2.02.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z68887; CAA93114.1; .
DR InterPro: IPR002937; Amino_oxidase.
DR InterPro: IPR000910; HMG_12_Box.
DR Pfam: PF01593; Amino_oxidase; 1.
DR SMART: SM00398; HMG; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 1028 1048 POTENTIAL.
SQ SEQUENCE 1273 AA; 142489 MW; 3FEFF6433DF3C0B6 CRC64;

alignment_scores:

Quality: 59.50 Length: 43
Ratio: 2.167 Gaps: 1
Percent Similarity: 62.791 Percent Identity: 34.884

alignment_block:

US-09-049-696-15 x YAR2_SCHPO ..

Align seg 1/1 to: YAR2_SCHPO from: 1 to: 1273

10 TTCATTGCTATTCAGGCTGTTGATAGGTCGATCTGAAATCAGAAATATC 59
287 PheMetProThrGlnLeuValSerAlaThrGluLeuSerIleValas 303
60 CAACATTGCACGAGTATCTTTGTTTATCTCTCCACAGACTCCGCCAGAGA 109
303 pAsnAlaVal.....LeuProIleProThrThrAlaProAlaVal 317
110 CACCTAGTCTGTGATGAACGTCGCTCCT 138
317 alValSerProProAlaSerSerPhePro 326

seq_name: SwissProt_39:HXA3_MOUSE

seq_documentation_block:

ID HXA3_MOUSE STANDARD; PRT; 443 AA.
AC P02831; Q61197;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HOMEOBOX PROTEIN HOX-A3 (HOX-1.5) (MO-10).
GN HOXA3 OR HOXA-3 OR HOX-1.5.
OS Mus musculus (Mouse).

This Page Blank (uspto)

Align seg 1/1 to: I51172 from: 1 to: 284

43 CTGAATCAGAAATCAACATGTCACGAGTATCTTTGTTATTCCTCC 92
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
145 LeuProSerGlnIleAlaThrCysAlaGlnThrThrMetSerLeuGlu 161
93 ACAGACTCCGCCAGACACCTAGTCTCT...GATGAACGTCGTCTCTT 139
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
161 nProThrProThrSerProGluProCysSerAsnThrSerSerAlaC 178

140 GTCCTAAT 147

|||||

178 ysProSer 180

seq_name: pir2:T10799

seq_documentation_block:

sucrose alpha-glucosidase (EC 3.2.1.48) / oligo-1,6-glucosidase (EC 3.2.1.10) - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C:Accession: T10799

R:Chandrasena, G.; Osterholm, D.E.; Sunitha, I.; Henning, S.J.

Gene 150, 355-360, 1994

A:Title: Cloning and sequencing of a full-length rat sucrose-isomaltase-encoding cDNA.

A:Reference number: 217155; MUID:95121929

A:Accession: T10799

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1841 <CHA>

A:Cross-references: EMBL:L25926; NID:g414818; PIDN:AAA65097.1; PID:g773669

A:Experimental source: strain Sprague-Dawley, intestine

C:Genetics:

A:Gene: SI

C:Function:

A:Description: sucrose alpha-glucosidase catalyzes the hydrolysis of the 1,2-alpha-D-glu

A:Pathway: carbohydrate digestion

C:Superfamily: sucrose/isomaltase; sucrose/isomaltase homology; trefoil homology

C:Keywords: carbohydrate digestion; glycosidase; hydrolase

F:199-846/Domain: sucrose/isomaltase homology <SIM>

F:937-983/Domain: trefoil homology <TRF>

alignment_scores:

Quality: 64.50 Length: 55

Ratio: 2.150 Gaps: 1

Percent Similarity: 54.545 Percent Identity: 32.727

alignment_block:

US-09-049-696-15 x T10799 ..

Align seg 1/1 to: T10799 from: 1 to: 1841

7 CTTTTCATTGCTATTTCAGGCTGTT..... 30

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

16 LeuPheIleIleValThrAlaIleAlaIleAlaLeuValThrValLeuAl 32

31 .GATAGTCGATCGATCAATGAGAATATCCAACTTCACGAGTATCTT 79

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

32 aThrLysValProAlaValGluGluIleLysSerProThrProThrSera 49

80 TCTTTATTCCTCCACAGACTCCGCCAGACACCTAGTCTCTGATGAACG 129

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

49 snSerThrProThrSerThrProThrSerThrSerThrProThrSerThr 65

130 TCTGCTCCTGTCTCT 144

||||| ||||| |||||

66 SerThrProSerPro 70

seq_name: pir2:C81058

seq_documentation_block:

hypothetical protein NMB1644 [imported] - Neisseria meningitidis (strain MC58 serogroup

C:Species: Neisseria meningitidis

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C:Accession: C81058

R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.

ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755

A:Accession: C81058

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-446 <TET>

A:Cross-references: GB:AE002515; GB:AE002098; NID:g7226894; PIDN:AAF41993.1; PID:g722

A:Experimental source: serogroup B, strain MC58

C:Genetics:

A:Gene: NMB1644

C:Superfamily: Neisseria meningitidis hypothetical protein NMB1644

alignment_scores:

Quality: 64.00 Length: 47

Ratio: 2.286 Gaps: 2

Percent Similarity: 59.574 Percent Identity: 40.426

alignment_block:

US-09-049-696-15/rev x C81058 ..

Align seg 1/1 to: C81058 from: 1 to: 446

143 ATATTAGACAGGAGCAGACGCTT.....TCATCAGGACT 115

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

314 LeuIleGlyLeuGlyLeuAspIleAlaThrLeuGlyGlySerLeuGlyLeu 330

114 AGGTGTCCTCT...GCCGAGTCTGTGGAGGAGTAATAACAAGATCTCGTG 68

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

330 uGlyThrAlaIleGlyGlyPheLeuGlyGlyIleLeuProAsnThrArgT 347

67 CAATGTCGATATTTCTGATTTTCAGATCGACCTTATCAACA 27

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

347 hrIleSerAspLysLeuAlaGlyArgGlnThrLeuHisThr 360

seq_name: pir2:B81817

seq_documentation_block:

probable integral membrane protein NMA1898 [imported] - Neisseria meningitidis (strai

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C:Accession: B81817

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo

; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Ouail, M.A.; Rajandre

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491

A:Reference number: A81775; MUID:20222556

A:Accession: B81817

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-464 <PAR>

A:Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB85119.1; PID:g738

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: NMA1898

C:Superfamily: Neisseria meningitidis hypothetical protein NMB1644

alignment_scores:

Quality: 64.00 Length: 47

Ratio: 2.286 Gaps: 2

Percent Similarity: 59.574 Percent Identity: 40.426

alignment_block:

US-09-049-696-15/rev x B81817 ..

Align seg 1/1 to: B81817 from: 1 to: 464

149 ATATTAGCAAGACGACGAGT.....TCATCAGGACT 115
: : : : : ||| ||| : : : : :
314 LeuilegyleuglyLeuaspilleAlaThrLeuGlySerLeuGlyLe 330
114 AGGTGCTCT...GGCGGAGTCTGTGGAGGATAAACAAGATCTCTG 68
||| : : : : : ||| : : : : :
330 uGlyThrAlaileGlyGlyPheLeuGlyGlyIleLeuProAsnThrArg 347
: : : : : ||| : : : : :
67 CAATGTTGGATATCTTGATTTTCAGATCGACCTTATCAACA 27
: : : : : ||| : : : : :
347 hrileSeraspLysLeuAlaGlyArgGlnThrLeuHisThr 360

seq_name: pir2:T49298

seq_documentation_block:

hypothetical protein T16L24.120 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000

C:Accession: T49298

R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z25017

A:Accession: T49298

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-549 <DAN>

A:Cross-references: EMBL:AL138659; GSPDB:GN00061; ATSP:T16L24.120

A:Experimental source: cultivar Columbia; BAC clone T16L24

C:Genetics:

A:Gene: ATSP:T16L24.120

A:Map position: 3

A:Introns: 6/2; 32/3; 71/3; 104/3; 139/2; 393/1; 434/1; 463/3

alignment_scores:

Quality: 64.00 Length: 30
Ratio: 2.909 Gaps: 0
Percent Similarity: 73.333 Percent Identity: 46.667

alignment_block:

US-09-049-696-15 x T49298 ..

Align seg 1/1 to: T49298 from: 1 to: 549

31 GATAAGTCGATCTGAATATCAGAAATATCCACATTCCACGAGTATCTTT 80
: : : : : ||| ||| : : : : :
223 GluservalaspLeuValSerProAspSerThrIleTyrgluThrSerSe 239
81 GTTATTCTCCACAGACTCCGCCAGACACCTAGTCTT 120
||| : : : : : ||| : : : : :
239 rPheValProSerSerProTyrglyPheProSerPro 252

seq_name: pir2:F86387

seq_documentation_block:

probable Pto kinase interactor [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: F86387

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

ansen, N.F.; Hughes, B.; Hulzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, X.H.; Liu, X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: F86387

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-760 <STO>
A:Cross-references: GB:AE005172; NID:gl1079512; PIDN:AAG29223.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

alignment_scores:

Quality: 64.00 Length: 19
Ratio: 4.571 Gaps: 0
Percent Similarity: 73.684 Percent Identity: 52.632

alignment_block:

US-09-049-696-15 x F86387 ..

Align seg 1/1 to: F86387 from: 1 to: 760

88 CTTCCACAGACTCCGCCAGACACCTAGTCTGATGAACGCTGCTCC 137
||| : : : : : ||| : : : : :
150 ProProgluSerProProSerLeuProAlaProAspProProSerAsnPr 166
138 TTGTCTCT 144
|||
166 OLeuPro 168

seq_name: pir2:I49140

seq_documentation_block:

p62 ras-GAP associated phosphoprotein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999

C:Accession: I49140

R:Richard, S.; Yu, D.; Blumer, K.J.; Hausladen, D.; Olszowy, M.W.; Connelly, P.A.; Sh

Mol. Cell. Biol. 15, 186-197, 1995

A:Title: Association of p62, a multifunctional SH2- and SH3-domain-binding protein, w

A:Reference number: I49140; MUID:95097990

A:Accession: I49140

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-443 <RES>

A:Cross-references: EMBL:U17046; NID:9608527; PIDN:AAA64997.1; PID:9608528

C:Keywords: phosphoprotein

alignment_scores:

Quality: 63.50 Length: 22
Ratio: 4.233 Gaps: 1
Percent Similarity: 68.182 Percent Identity: 59.091

alignment_block:

US-09-049-696-15 x I49140 ..

Align seg 1/1 to: I49140 from: 1 to: 443

88 CTTCCACAGACTCCGCCAGACACCTAGTCTGATGAACG..... 129
||| : : : : : ||| : : : : :
64 ProProLeuLeuProProSerThrProGlyProAspAlaThrValValGl 80
130 .TCTGCTCTCTGCTCT 144
||| : : : : : |||
80 ySerAlaProThrPro 85

seq_name: pir2:T34342

seq_documentation_block:

hypothetical protein T08H4.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34342

R:Stelljes, L.

submitted to the EMBL Data Library, October 1995

A:Description: The sequence of C. elegans cosmid T08H4.

A:Reference number: Z21510
 A:Accession: T34342
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-862 <STE>
 A:Cross-references: EMBL:U39470; PIDN:AAC71118.1; GSPDB:GN00020; CESP:T08H4.1
 A:Experimental source: strain Bristol N2; clone T08H4
 C:Genetics:
 A:Gene: CESP:T08H4.1
 A:Map position: 2
 A:Introns: 17/3; 44/2; 170/2; 259/3; 291/3; 553/1; 768/3; 804/3; 860/2

alignment_scores:
 Quality: 63.00 Length: 48
 Ratio: 2.100 Gaps: 3
 Percent Similarity: 62.500 Percent Identity: 39.583

alignment_block:

US-09-049-696-15 x T34342 ..

Align seg 1/1 to: T34342 from: 1 to: 862

```

7 CTTTTCATTGCTATTCAGGCTGTTGATAAGTCGATCTGAATCAGAAAT 56
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
792 LeuSerIleSerLeuGluAlaProThrValLeuGluLysAspGluAs 808

57 ATCCACACATTGCACGAGTATCTTTGTTTATTCCTCCACAGACTCCGCCAG 106
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
808 pGlnHisSerIleGluValSer.....ValProPro.....ProProG 821

107 AGACACCTAGTCCTGATGAACGCTGCTCTCTGCTTAATATT 150
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
821 luLeuProProSerLysThr.....ProSerProAsnLeu 833
  
```

seq_name: pir2:S09269

seq_documentation_block:

Ig alpha chain C region - rabbit (fragment)
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 16-Jul-1999
 C:Accession: S09269
 R:Burnett, R.C.; Hanly, W.C.; Zhai, S.K.; Knight, K.L.
 EMBL J. 8, 4041-4047, 1989
 A:Title: The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13
 A:Reference number: S09264; MUID:90076124
 A:Accession: S09269
 A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-357 <BUR>
 C:Superfamily: immunoglobulin C region; immunoglobulin homology
 C:Keywords: immunoglobulin
 F:142-208/Domain: immunoglobulin homology <IMM>

alignment_scores:
 Quality: 62.50 Length: 20
 Ratio: 4.167 Gaps: 1
 Percent Similarity: 75.000 Percent Identity: 55.000

alignment_block:

US-09-049-696-15 x S09269 ..

Align seg 1/1 to: S09269 from: 1 to: 357

```

88 COTCCACAGACTCCGCCAGACACCTAGTCCT...GATGAACGCTGCC 134
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99 ProProGluCysGlnProProThrProGlyProSerAspThrThrCy 115

135 TCCTTGTCCT 144
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
115 sProCysPro 118
  
```

seq_name: pir2:T19736
 seq_documentation_block:
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1181 <WIL>
 A:Cross-references: EMBL:281479; PIDN:CAB03946.1; GSPDB:GN00028; CESP:C34F6.9
 A:Experimental source: clone C34F6
 C:Genetics:
 A:Gene: CESP:C34F6.9
 A:Map position: X
 A:Introns: 121/1; 242/1; 627/1; 655/1; 724/1; 757/3; 837/3; 939/1; 986/1; 1036/2; 112

alignment_scores:

Quality: 62.50 Length: 42
 Ratio: 2.155 Gaps: 1
 Percent Similarity: 69.048 Percent Identity: 38.095

alignment_block:

US-09-049-696-15 x T19736 ..

Align seg 1/1 to: T19736 from: 1 to: 1181

```

101 GATCTTTTCATTGCTATTCAGGCTGTTGATAAGTCGATCTGAATCAGA 53
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
689 AspIleGluGlnAlaValGlnAlaIleAspAsnValHisLysAlaLys 705

54 AATATCCACATTCACGAGTATCTTTTATTCCTCCACAG...ACTC 100
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
705 sArgGluArgLysSerArgSerSerIleThrValSerGluAspGlyTyrP 722

101 GCGCAGACAGACCTAGTCCTGATGAA 126
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
722 roProGluThrValSerProGlnGlu 730
  
```

seq_name: pir2:A54659

seq_documentation_block:

DNA repair protein Xrcc-1 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 05-Nov-1999
 C:Accession: A54659
 R:Brookman, K.W.; Tebbbs, R.S.; Allen, S.A.; Tucker, J.D.; Swiger, R.R.; Lamerdin, J.E
 Genomics 22, 180-188, 1994
 A:Title: Isolation and characterization of mouse Xrcc-1, a DNA repair gene affecting
 A:Reference number: A54659; MUID:95048367
 A:Accession: A54659
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-631 <BRO>
 A:Cross-references: GB:U02887; NID:g806691; PIDN:AAA93115.1; PID:g409953
 C:Genetics:
 A:Gene: Xrcc-1
 C:Keywords: DNA repair

alignment_scores:

Quality: 62.00 Length: 18
 Ratio: 4.133 Gaps: 0
 Percent Similarity: 83.333 Percent Identity: 55.556

alignment_block:

US-09-049-696-15 x A54659 ..

Align seg 1/1 to: A54659 from: 1 to: 631

```
91 CCAAGACTCCGCCAGAGACACCTAGTCCTGATGAAACGTCCTCCTTG 140
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
444 ProSerSerProProArgProProThrProLysGluThrLysAlaProSe 460
   |||
141 TCCT 144
   |||
460 rPro 461
```

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	4759	100.0	9	914	4	Q9UNF6	Q9unf6 homo sapien
2	4756	99.9	9	914	4	Q95151	Q95151 homo sapien
3	4751	98.8	9	914	4	Q9UPC6	Q9upc6 homo sapien
4	3789.5	79.6	917	6	Q9FUB5	Q9fub5 sus scrofa	
5	3656.5	76.8	913	11	Q88826	Q88826 mus muscul	
6	3647.5	76.6	913	11	Q9D726	Q9d726 mus muscul	
7	2872.5	60.4	917	4	Q9UNF7	Q9unf7 homo sapien	
8	2361.5	49.6	902	11	Q9OX15	Q9ox15 mus muscul	
9	2358.5	49.6	902	11	Q9E0R4	Q9eqr4 mus muscul	
10	2353.5	49.5	902	11	Q9R070	Q9r070 mus muscul	
11	2351	49.4	901	11	Q88860	Q88860 mus muscul	
12	2334	49.0	905	6	O18741	O18741 bos taurus	
13	2131	44.8	820	6	O18743	O18743 bos taurus	
14	2128	44.7	794	6	O18742	O18742 bos taurus	
15	1998.5	42.0	943	4	Q9Y6N2	Q9y6n2 homo sapien	
16	1989.5	41.8	943	4	Q9UCQ9	Q9ucq9 homo sapien	
17	1423	29.9	469	4	Q9NXP1	Q9npx1 homo sapien	
18	947.5	19.9	342	6	O18744	O18744 bos taurus	
19	749.5	15.7	262	4	Q9Y6N3	Q9y6n3 homo sapien	

QY	1	MGPFKSSVFILILHLLLEGALSNLSLIQLNNGYBGIIVVADPNVPEDETLIQQIKDMVTOA	60
..
..
DB	1	MGPFKSSVFILILHLLLEGALSNLSLIQLNNGYBGIIVVADPNVPEDETLIQQIKDMVTOA	60
..
..
QY	61	SLYLFEATGKRRFYFKNVAIILPETWTKTADYVRPKLETYKNADVLVAESTPGCNDEPYTE	120
..
..
DB	61	SLYLFEATGKRRFYFKNVAIILPETWTKTADYVRPKLETYKNADVLVAESTPGCNDEPYTE	120
..
..

QY	121	QMGNCCEKGERIHLTPDFTAGKKLABYGPQGRAFVHEWAHLRWGVDFEYNNDEKPYLSNG	180
Db	121	QMGNCCEKGERIHLTPDFTAGKKLABYGPQGRAFVHEWAHLRWGVDFEYNNDEKPYLSNG	180
QY	181	RIQAVRCSAGITGTNNVKKCQGGSCYTKKCTFNKVTGLYEKCECFVLOSROTEKASIMFA	240
Db	181	RIQAVRCSAGITGTNNVKKCQGGSCYTKKCTFNKVTGLYEKCECFVLOSROTEKASIMFA	240
QY	241	QHVDISIVEFCTEQHNHKEAPNKONKCNLRSTWEVIRDSEDFKKTPTMTQPPNPTFSLL	300
Db	241	QHVDISIVEFCTEQHNHKEAPNKONKCNLRSTWEVIRDSEDFKKTPTMTQPPNPTFSLL	300
QY	301	QIGORIVCLVLDKSGSMATGNRLNRUNQAQGLFLQTVELGWSGVMTFDSAHHVQSELI	360
Db	301	QIGORIVCLVLDKSGSMATGNRLNRUNQAQGLFLQTVELGWSGVMTFDSAHHVQSELI	360
QY	361	QINSGSDROTAKRLPAAASGGTSTCSGLRSFTVIRKKYPTDGSSEIVLLTGDGENTISG	420
Db	361	QINSGSDROTAKRLPAAASGGTSTCSGLRSFTVIRKKYPTDGSSEIVLLTGDGENTISG	420
QY	421	CFNEVKQSGAIHTVALGPSAAQELBELSKMTGGLOTYASDOVQNGLLDAFGALLSSNG	480
Db	421	CFNEVKQSGAIHTVALGPSAAQELBELSKMTGGLOTYASDOVQNGLLDAFGALLSSNG	480
QY	481	AVSORSTQLESKGLTLQNSQWNGTVIVDSTVGKDTFLITWTTPQPQILLMDPSGQKQG	540
Db	481	AVSORSTQLESKGLTLQNSQWNGTVIVDSTVGKDTFLITWTTPQPQILLMDPSGQKQG	540
QY	541	GFVVDKNTMAYLIQPGIAKVGTWKYSLOASSOTLTLVTVSRASNATLPPIVTVSKTNKD	600
Db	541	GFVVDKNTMAYLIQPGIAKVGTWKYSLOASSOTLTLVTVSRASNATLPPIVTVSKTNKD	600
QY	601	TSKFPSPLVVIYANIROGASPIILRASVTALIESVNGKVTLELLDNGAGADATKDDGVYSR	660
Db	601	TSKFPSPLVVIYANIROGASPIILRASVTALIESVNGKVTLELLDNGAGADATKDDGVYSR	660
QY	661	YFTYDTNGRYSVKVRALGVNAARRVVIPOQSGALYIPGWIENDEIQWNPPRPEINKDD	720
Db	661	YFTYDTNGRYSVKVRALGVNAARRVVIPOQSGALYIPGWIENDEIQWNPPRPEINKDD	720
QY	721	VOHKQVCFSTSSGGSFVASDVPNAPIDLFPPGQITDLKAEIHGGSLLNLWTAPGDDY	780
Db	721	VOHKQVCFSTSSGGSFVASDVPNAPIDLFPPGQITDLKAEIHGGSLLNLWTAPGDDY	780
QY	781	DHGTAKHYIIRISTSLDRKFENESLQVNTTALIPKEANSEVEFLFKPENITTFENGIDL	840
Db	781	DHGTAKHYIIRISTSLDRKFENESLQVNTTALIPKEANSEVEFLFKPENITTFENGIDL	840
QY	841	FIAIQAVDKVDLKSEISNIARVSLFPPOTPETSPDETSAPCPNIHINSTIPGHIHLK	900
Db	841	FIAIQAVDKVDLKSEISNIARVSLFPPOTPETSPDETSAPCPNIHINSTIPGHIHLK	900
QY	901	IMMKWIGELQLSLTA 914	
Db	901	IMMKWIGELQLSLTA 914	
RESULT	2		
Q95151			
ID	O95151	PRELIMINARY; PRT; 914 AA.	
AC	O95151;		
DT	01-MAY-1999 (TrEMBLrel. 10, Created)		
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)		
DE	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DE	CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.		
GN	HCICAI.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		

```

QY 781 DHGTAHKYIIRISTSLDLRDKNESLOVNTTALIPKEANSEEVFLFKPENITFNGTDL 840
Db 781 DHGTAHKYIIRISTSLDLRDKNESLOVNTTALIPKEANSEEVFLFKPENITFNGTDL 840
QY 841 FTAIAQAVDKVLKSEISNIARVSLFPPTPTPSPDETSPACPNIHINSTITPGHIHLK 900
Db 841 FTAIAQAVDKVLKSEISNIARVSLFPPTPTPSPDETSPACPNIHINSTITPGHIHLK 900
QY 901 IMWKWIGELQSLA 914
Db 901 IMWKWIGELQSLA 914

RESULT 3
Q9UPC6 PRELIMINARY; PRT; 914 AA.
AC Q9UPC6;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-L1.
GN HCLCAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Elble R.C., Ji H.L., Schreier K.D., Fuller C.M.,
RA Pauli B.U.
RT "Genomic cloning, molecular characterization, and functional analysis
RT of human HCLAL, the first human member of the family of Ca2+-activated
RT Cl-channel proteins."
RL Genomics 54:200-214(1998).
DR EMBL; AF039401; AAC95429.1; -.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; WFEA_gamma.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; WFEA; 1.
DR SMART; SM00327; WFA; 1.
SQ SEQUENCE 914 AA; 100160 MW; 7761362B66752AA3 CRC64;

Query Match 99.8%; Score 4751; DB 4; Length 914;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 912; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGPFKSSVFILHLLEGALSLSLIQNNNGYEGIVVAIDPNVPDETLLIQIKDMVTOA 60
Db 1 MGPFKSSVFILHLLEGALSLSLIQNNNGYEGIVVAIDPNVPDETLLIQIKDMVTOA 60
QY 61 SLYLFEATGKRFYKFNVAIIPETWTKADYVRPKLETYKNADVLVAESTPPGNDEPYTE 120
Db 61 SLYLFEATGKRFYKFNVAIIPETWTKADYVRPKLETYKNADVLVAESTPPGNDEPYTE 120
QY 121 QMNCCEKERIHLTPDFTAGKKLAEGYPOGAFVHEWAHLRWGVDEYNNDKEFVLSNG 180
Db 121 QMNCCEKERIHLTPDFTAGKKLAEGYPOGAFVHEWAHLRWGVDEYNNDKEFVLSNG 180
QY 181 RQAVRCSAGITGTVNVKKCGGSCYTKRCTNKNVTLGVEKGEFVLQSRQTEKASIMFA 240
Db 181 RQAVRCSAGITGTVNVKKCGGSCYTKRCTNKNVTLGVEKGEFVLQSRQTEKASIMFA 240
QY 241 QHVDSTVEFCTEQNHKNEAPNKNOKCNLRSTWEIRDSDEFKKTPMTTQPPNPFTSL 300
Db 241 QHVDSTVEFCTEQNHKNEAPNKNOKCNLRSTWEIRDSDEFKKTPMTTQPPNPFTSL 300
QY 301 QICQRTVCLVLDKSGSMATGNRLNRLNQAGQLFLLOTVELGSGWGVMTFDSAHHVSELI 360
Db 301 QICQRTVCLVLDKSGSMATGNRLNRLNQAGQLFLLOTVELGSGWGVMTFDSAHHVSELI 360
QY 361 QINSGSDRDTLAKRLPAAASGGTSCSGLRSFAFTVIRKKYPTDGSSEIVLLTDGEDNTISG 420

```

```

Db 361 QINSGSDRDTLAKRLPAAASGGTSCSGLRSFAFTVIRKKYPTDGSSEIVLLTDGEDNTISG 420
QY 421 CFNEVQSGAI IHTVALGPSAAQEELELSKMTGGLQTYASDQVQNNGLIDAFGALSNG 480
Db 421 CFNEVQSGAI IHTVALGPSAAQEELELSKMTGGLQTYASDQVQNNGLIDAFGALSNG 480
QY 481 AVSQRISQLESKGLTLQNSOMMNGTVIVDSVGVKDTLFLITWTTPQPIILLWDPGKQK 540
Db 481 AVSQRISQLESKGLTLQNSOMMNGTVIVDSVGVKDTLFLITWTTPQPIILLWDPGKQK 540
QY 541 GFVVDKNTKMYLQIPGIARVGTWKYSLOASSOTLTLTTSRASNATLPPITVTSKTKND 600
Db 541 GFVVDKNTKMYLQIPGIARVGTWKYSLOASSOTLTLTTSRASNATLPPITVTSKTKND 600
QY 601 TSKFSPPLVYVNIROGASPILRASVTALIESVNGKTVTLLELDNGAGADATKDDGVYSR 660
Db 601 TSKFSPPLVYVNIROGASPILRASVTALIESVNGKTVTLLELDNGAGADATKDDGVYSR 660
QY 661 YFTTYDTNGRYSVKVVGALGVNAARRVIPPQSGALYIPGWIEDEIQWNPPEINKDD 720
Db 661 YFTTYDTNGRYSVKVVGALGVNAARRVIPPQSGALYIPGWIEDEIQWNPPEINKDD 720
QY 721 VOHKQVCFESRTSSGSEFVSDVNPAPIPDLFPPQITDLKAEIHGGSLLNLTWAPGDDY 780
Db 721 VOHKQVCFESRTSSGSEFVSDVNPAPIPDLFPPQITDLKAEIHGGSLLNLTWAPGDDY 780
QY 781 DHGTAHKYIIRISTSLDLRDKNESLOVNTTALIPKEANSEEVFLFKPENITFNGTDL 840
Db 781 DHGTAHKYIIRISTSLDLRDKNESLOVNTTALIPKEANSEEVFLFKPENITFNGTDL 840
QY 841 FTAIAQAVDKVLKSEISNIARVSLFPPTPTPSPDETSPACPNIHINSTITPGHIHLK 900
Db 841 FTAIAQAVDKVLKSEISNIARVSLFPPTPTPSPDETSPACPNIHINSTITPGHIHLK 900
QY 901 IMWKWIGELQSLA 914
Db 901 IMWKWIGELQSLA 914

RESULT 4
Q9TUB5 PRELIMINARY; PRT; 917 AA.
AC Q9TUB5;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN.
GN AECC.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ILEAL MUCOSA;
RA Gaspar K.J., Gabriel S.E., Racette K.J., Forsyth G.W.;
RT "A cDNA involved in porcine exocrine chloride conductance.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF095584; AAF00077.1; -.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; WFEA_gamma.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; WFEA; 1.
DR SMART; SM00327; WFA; 1.
SQ SEQUENCE 917 AA; 100735 MW; 649F41944F7EAD19 CRC64;

Query Match 79.6%; Score 3789.5; DB 6; Length 917;
Best Local Similarity 78.6%; Pred. No. 1.6e-262;
Matches 723; Conservative 81; Mismatches 107; Indels 9; Gaps 3;

QY 1 MGPFKSSVFILHLLEGALSLSLIQNNNGYEGIVVAIDPNVPDETLLIQIKDMVTOA 60

```

Db	1	MGSFSSFLVHLHLEGAQSNLSIQLNGNGYEGIVIAIDPNVPEDERLIQIKDMVTKA	60
Qy	61	SLYLFEATGRFRYFKNVAILIPETWTKADYVVRPKLETYNADVLVAESTPPGNDPEYTE	120
Db	61	SPYLFEATEKFRYFKNVAILIPASWKAPEYVVRPKLETYNADVLVAESTPPGNDPEYTE	120
Qy	121	OMGNGCKEGERIHLTPDFIAGKKIAEYGPQGRAFVHWAHLRWGVDFEYNNDEKFFYLSNG	180
Db	121	OMGNGCKEGERIHLTPDFVAGKKVLYQPGQGRVFWHWAHLRWGVDFEYNNDEKFFYLSNK	180
Qy	181	RIQAVRCSAGITGNVVKKOGGSCYTKRFTNFVGLYKGECEFFVLOSQTEKASTMFA	240
Db	181	KQPVICSAIRGNTVLPKOGGSCVTKPCRADRVTLGFOKECEFFIPDPOOSEKASTMFA	240
Qy	241	QHVDSIVFECTEONHNKEAPKQKCNLRSTWVIRDSDFKTKTPMTTPPPNPTFSL	300
Db	241	QSIDTVVFECKEKNHKEAPNDQKCNLRSTWVIRDSDFKTKTPMTTPPPNPTFSL	300
Qy	301	QIGQIRIVCLVLDKSGMATGNRLNRLNOAGOLFILOTVELGSGWVMTFDSAAHVQSELI	360
Db	301	QIGQIRIVCLVLDKSGMTVGGRLKRLNOAGKLFLLQTVQEGAWGVMTFDSAAHVQSELY	360
Qy	361	QINSGSDRDTLAKRLPAAASGTSICSLRSATVIRKKYPTDGESEIVLLTDGEDNTISG	420
Db	361	QINSAERDALARSPTAASGTSICSLRSATVIRKKYPTDGESEIVLLTDGEDNTISA	420
Qy	421	CFNEVKOSGAIHTVALGPSAAQELSELSKMTGGLOTVASOVQNGNLIDAFGALSSNG	480
Db	421	CPPEVKONGAIHTVALGPSAAQELSELSKMTGGLOTVASOVAENGLIDAFGALSSNR	480
Qy	481	AVSORSIOLESGLTLONSQWNGTVIVDSVTGKDTFLITWTTO--PPQILLMDPSQOK	538
Db	481	ASORSIOLESGLTLONNENWNGTVIVDSVTGKDTFLITLTLERKFLSPIPEFGVPSGRS	540
Qy	539	QCGFVVDKNTKMAIYLOIPGIAKVTGWTYSLQASSOTLTLTVTSRASNLTPITVTSKN	598
Db	541	QDSFLVGKHKMAYFOVPGTAKVGMWYSLQASSOTLTLTVTSRRSSATLPVTVTSKMN	600
Qy	599	KDTSKPPSLVYVANTROGASPLTRASVTALTESVNGKTVTLLELDNGAGADATKDDGY	658
Db	601	KDTSKPPSLVYVANTKHOGLTLPILRAKVTALTESVNGKTVTLLELDNGAGADATKNDGY	660
Qy	659	SRVFTTYDFTNGRYSVKVRALGGVNAARRVYPOOSGALYIPGHIENDEIOWNPPRPEINK	718
Db	661	SRVFTAYDANGRYSVKVRALGGVNTPRRRAPPLWAGMYIRGWIENGELKWNPPRPEINK	720
Qy	719	DDVQHKOVCFSSSGGSFVASDPNAPIDLPFPQGITDLKAEIHGGSILNLTWTAPGD	778
Db	721	DDVQHKOVCFSSSGGSFVASDPKSPIDLPFPCKITDLKAGIQDNLNLTWTAPGD	780
Qy	779	DYDHGTAKYIIRISTSLDLRKNESLOVNTTALIPKEANSEEVFLKPEINITENGT	838
Db	781	DYDHGRADRYIIRISTNLDRKFNDSVQVNTDILIPKEANSEEVFLKPEGIPFNGT	840
Qy	839	DLFTAQVADKVLKSEISNTARYSLFI---PPQTPPETPSPDETSAFCNPIHINSTIP	894
Db	841	DLFTAQVADKVLKSEISNTARYSLFI---PPQTPPETPSPDETSAFCNPIHINSTIP	897
Qy	895	GIHILKIMWKGELQLSIA 914	
Db	898	GIHILKIMWKGELQLSIA 917	
RESULT	5		
O88826			
ID	O88826	PRELIMINARY; PRT; 913 AA.	
AC	O88826;		
DT	01-NOV-1998 (TremBLrel. 08, Created)		
DT	01-NOV-1998 (TremBLrel. 08, Last sequence update)		
DT	01-JUN-2001 (TremBLrel. 17, Last annotation update)		
DE	G0B-5 PROTEIN.		
GN	CLCA3 OR G0B-5.		

OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
QX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=INTESTINE;		
RA	Komiyama T., Tanigawa Y., Hirohashi S.;		
RT	"Cloning of a gene, gob-5, which is expressed in intestinal goblet		
RT	cells in mice."		
RL	Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB017156; BAA33743.1; -.		
DR	MGI; MGI:1346342; Clca3.		
DR	InterPro; IPR00131; ATPase_gamma.		
DR	InterPro; IPR002035; VMFA.		
DR	SMART; SM00327; VWA; 1.		
DR	PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.		
DR	PROSITE; PS0234; VMFA; 1.		
SQ	SEQUENCE 913 AA; 10070 MW; A7FA2F9E1089806D CRC64;		
	Query Match 76.8%; Score 3656.5; DB 11; Length 913;		
	Best Local Similarity 75.8%; Pred. No. 5.3e-253;		
	Matches 694; Conservative 102; Mismatches 112; Indels 7; Gaps 4;		
Qy	1	MGPFKSVFLLIHLLEGALSLSLIOLNNGYEGIVIAIDPNVPEDETLIQIKDMVTKA	60
Db	1	MESLKSPVFLIHLLEGVLSLSLIOLNNGYEGIVIAIDHDVPEDEALIQIKDMVTKA	60
Qy	61	SLYLFEATGRFRYFKNVAILIPETWTKADYVVRPKLETYNADVLVAESTPPGNDPEYTE	120
Db	61	SPYLFEATGRFRYFKNVAILIPESWKAPEYVVRPKLETYNADVLVAESTPPGNDPEYTE	120
Qy	121	OMGNGCKEGERIHLTPDFIAGKKIAEYGPQGRAFVHWAHLRWGVDFEYNNDEKFFYLSNG	180
Db	121	HIGAGGKGIHLTPDFIAGKKIATQYQPDRTFVHWAHLRWGVDFEYNNDEKFFYLSNG	180
Qy	181	RIQAVRCSAGITGNVVKKOGGSCYTK-RCTFNKVTGLYKGECEFFVLOSQTEKASIMF	239
Db	181	KQAVRCSAATGNVVRRCGGSCITNGKVCIDRVTLGKNCVFVDPDHPHOKASIMF	240
Qy	240	AQVDSIVFECTEONHNKEAPKQKCNLRSTWVIRDSDFKTKTPMTTPPPNPTFSL	299
Db	241	NONINSVFECTEONHNKEAPNDQKCNLRSTWVIRDSDFKTKTPMTTPPPNPTFSL	300
Qy	300	LOIGQIRIVCLVLDKSGMATGNRLNRLNOAGOLFILOTVELGSGWVMTFDSAAHVQSEL	359
Db	301	LOIGQIRIVCLVLDKSGMLNDRNRMNQASRLFLQTVQEGSWVMTFDSAAHVQSEL	360
Qy	360	IQINSGSDRDTLAKRLPAAASGTSICSLRSATVIRKKYPTDGESEIVLLTDGEDNTIS	419
Db	361	KOLNSGADRDLLIHLPTVSAGGTSICSLRSATVIRKKYPTDGESEIVLLTDGEDNTIS	420
Qy	420	GCNFVKOSGAIHTVALGPSAAQELSELSKMTGGLOTVASOVQNGNLIDAFGALSSGN	479
Db	421	SCFDLVKOSGAIHTVALGPSAAQELSELSKMTGGLOTVASOVQNGNLIDAFGALSSGN	480
Qy	480	GAVSORSIOLESKGLTLONSQWNGTVIVDSVTGKDTFLITWTTOPOQILLMDPSQOK	539
Db	481	AAIAQHSIOLESRGVNLQNNQWNGSVIVDSVSGKDTFLITWTTHPTTIFWIDPSGVEQ	540
Qy	540	GGFVVDKNTKMAIYLOIPGIAKVTGWTYSLQASSOTLTLTVTSRASNLTPITVTSKN	599
Db	541	NGFILDITTKVAYLQVPGTAKVGMWYSLQASSOTLTLTVTSRASNLTPITVTSKN	600
Qy	600	DTSKPPSLVYVANTROGASPLTRASVTALTESVNGKTVTLLELDNGAGADATKDDGY	659
Db	601	NTGKFPSPVTVASIROGASPLTRASVTALTESVNGKTVTLLELDNGAGADATKDDGY	660
Qy	660	RYFTTDTNGRYSVKVRALGGVNAARRVYPOOSGALYIPGHIENDEIOWNPPRPEINK	719
Db	661	RYFTAFDANGRYSVKVRALGGVNTDRQRAAPPKRNAMYIDGWIEDGEVWRNPPRPEIN	719


```

QY 720 DVQHKQVCFSTSSGGSFVASDVP-NAPIDPLFPQGITDLKAEIHGSLINLTWAPGD 778
DB 720 -VDQKLCFSRTSSGGSFVATNPAAAPIDLPFPQGITDLKASIQGNLNLWTAPGD 778
QY 779 DYDHGTAKHYIIRISTSIDLRKFNESLQVNTALIPKEANSEEVFLFKPENITFNGT 838
DB 779 DYDHGRASNVIIRMTSIVDLRHFNTSLQVNTGLIPKEASSEIEFEFLGCGTNGT 838
QY 839 DLFIATQAVDKVLDKSEISNIARVSLFPPQTPPETSPDTSAPCPNIHINSTPIGHI 898
DB 839 DIFIATQAVDKSLKSEISNIARVSFIPAQEP---PIPEDSTPPCPDISINSTPIGHI 895
QY 899 LKIMWKWIGELQSLI 913
DB 896 LKIMWKWLGEMQVTL 910

RESULT 6
Q9D726 PRELIMINARY; PRT; 913 AA.
ID Q9D726
AC Q9D726;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE CLORIDE CHANNEL CALCIUM ACTIVATED 3.
GN CLCA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=STOMACH;
RX MEDLINE=21085660; PubMed=11217851;
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Frieschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo T., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner N., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK008659; BAB25815.1; -.
DR MGD: MGI:1346342; Clca3.
DR InterPro: IPR000131; ATPase_gamma.
DR InterPro: IPR002035; VWFA.
DR SMART: SM00327; VWFA; 1.
DR PROSITE: PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE: PS50234; VWFA; 1.
SQ SEQUENCE 913 AA; 100111 MW; 491E8584B06D9A89 CRC64;

```

```

Query Match      76.6%; Score 3647.5; DB 11; Length 913;
Best Local Similarity 75.6%; Pred. No. 2.3e-252;
Matches 692; Conservative 103; Mismatches 113; Indels 7; Gaps 4;

```

```

QY 1 MGPFKSSVILILHLEGALSLSLIQNNNGYEGIVADPNVPEDETLLQIQKDMVTA 60
DB 1 MESLKSVPVLLILHLEGVLSLSLIQNNNGYEGIVADHDVPEDEALIQHKDMVTA 60
QY 61 SLYLFEATGRFFKNAIVILPETWTKADYVRPKLETYNADVLVAESTPPGNDEPYTE 120

```

```

DB 361 SPYLFEATGRFFKNAIVILIPESWAKPEYTRKLETFKNADVLVSTTSLGNDEPYTE 120
QY 121 QMNGCGSGRIILTPDFIAGKLALEYGPOGRAFVHWAHLRWGVFDEYNNDEKFIYLSNG 180
DB 121 HICACGEGRIILTPDFLAGKLLTGPODRTFVHWAHFRWGVFDEYNNDEKFIYLSK 180
QY 181 RIOAVRCSAGITGVNVVKCGGSCYTK-RCTFNKVTGLYEKGCEFFVLSQSKTEKASIMF 239
DB 181 KPOAVRCSAATGKNQVRRCGGSCITNGKVIDRVTGLYKDCNCFVDPHQNEKASIMF 240
QY 240 AOHVDSIVFECTONHNKEAPNKNOKNCLNRSTHEVIRDSDEPKTKTTPMTTOPNPFTSL 299
DB 241 NQNTNSVVEFTEKNHNOEAPNDQNCNLRSTWETVIOESDFKQTPMTAQPPAPFTSL 300
QY 300 LOIQQRIVCLVLDKSGSMATGNRLNRLNQAGQLFLLQTVELGWSWGMVTFDSAHHVQSEL 359
DB 301 LOIQQRIVCLVLDKSGSMLNDRNRNMQASRLFLQTVEOGSWGMVTFDSAAYVQSEL 360
QY 360 IQNNGSDROTFLAKRLPAAASGGTSCISGLRSATVIRKKYPTDGCSEIVLLTDGEDNTIS 419
DB 361 KQLNSGADRLLIKHLPTVSGAGTSCISGLRTAFTVIRKKYPTDGCSEIVLLTDGEDNTIS 420
QY 420 GCFNEVKQSGAITHTVALGPSAAQELRELSKMTGLOTYASDOVQNNGLIDAFGALSNG 479
DB 421 SCFDLVKQSGAITHTVALGPAARKEQLSKMTGLOTYSSDVQNNGLVDFAALSSGN 480
QY 480 GAVSQRSIQLESKGLTLQNSQWNGTVIVDSTVGKDTLFLITWTTPQPIILLWDPGQK 539
DB 481 AATAQHSIQLESRGVNLQNNQWNGSVIVDSSVGKDTLFLITWTTPHTFIDWPSGVEQ 540
QY 540 GGFVVDKTKMAYLIQPIGIAKVGTWYSLQASSOTLTLTVTSRASNATLPPITVTSKTK 599
DB 541 NGFILDITTKVAVLQVPGTAKGVFWKYSIQASSOTLTLTVTSRAASATLPPITVTPVVK 600
QY 600 DTSKFPSPVYVYNIROGASPIRLASVTALIESVNGKTVTLELLDNGAGADATKDDGVYS 659
DB 601 NTKRFPSPVTVYASIROGASPIRLASVTALIESVNGKTVTLELLDNGAGADATKDDGVYS 660
QY 660 RYFTTYDTNGRYSYKVRALGVNAARRVPIQQSGALYIPQGIENDEIQMPPPEINKD 719
DB 661 REFTAFDANGRYSYKIALGVNDSRQRAAPPKNRPVYIDWIEDGEVRMPPRPETSY- 719
QY 720 DVQHKQVCFSTSSGGSFVASDVP-NAPIDPLFPQGITDLKAEIHGSLINLTWAPGD 778
DB 720 -VDQKLCFSRTSSGGSFVATNPAAAPIDLPFPQGITDLKASIQGNLNLWTAPGD 778
QY 779 DYDHGTAKHYIIRISTSIDLRKFNESLQVNTALIPKEANSEEVFLFKPENITFNGT 838
DB 779 DYDHGRASNVIIRMTSIVDLRHFNTSLQVNTGLIPKEASSEIEFEFLGCGTNGT 838
QY 839 DLFIATQAVDKVLDKSEISNIARVSLFPPQTPPETSPDTSAPCPNIHINSTPIGHI 898
DB 839 DIFIATQAVDKSLKSEISNIARVSFIPAQEP---PIPEDSTPPCPDISINSTPIGHI 895
QY 899 LKIMWKWIGELQSLI 913
DB 896 LKIMWKWLGEMQVTL 910

RESULT 7
Q9UNF7 PRELIMINARY; PRT; 917 AA.
ID Q9UNF7
AC Q9UNF7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 2.
GN CAC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;

```

[illegible]

```
Db 360 YLIKITSSDYOKITANLPQOASGGTICHGLOAGFOAITSSDOSTSGSEIVLLTDEGN 419
QY 417 TISGCNEVQSGAIHTHTVALGSAQAELSELSKMTGGLOQTYASDQVQNNGLIDAFGALS 476
Db 420 GIRSCEAVSRSGAIHTHTIALGSAARELETLSDMTGGGLRYFANKDL--NSLIDAFSRIS 477
QY 477 SGNCAVSRISOLSKGLTLNLSQWNCVIVDSTVGKDTPLFTITWTPPOILLWDPSC 536
Db 478 STSGVSQALQLESKAFDVRAGAWINGVPLDSTVGNDFVITWMVKPEIILQDPKG 537
QY 537 QK--QGFFVVDK--NTKMAVLIQPIGAKVGTWKYSLOAS--SQTLLTAVTSRASNATLPPI 592
Db 538 KKYITSDFOQDKLIRNSARLIQIPGTAETGTWTYSITGKSQLITMIVTTRASPTMEPLL 597
QY 593 VTSKTNKDTSKFSPPLVYVNIIRQASPIRLASVTALIESVNGKTVLELDNGAGADAT 652
Db 598 ATAHMSQSTAQYPSRMIVYARVSGQLPVLGAVNTALIEAHEGHQVTLLELDNGAGADTV 657
QY 653 KDGVSRYFTYDTHNGRYSVKVRALGGVNAARRV--IPQOSGALYIPGWLENDEIQW 709
Db 658 KNDGIYTRYFTYHNGRYSKLV--VOAQNKTRLSLRQNKSLYIPGYVENGKIVL 713
QY 710 NPPRPINKDDVQHKQVCFSTRSSGGSFVASDVPNAPIDP-----LPPGQITDLKAEIH 764
Db 714 NPPRPDQBEAIEATVEDFNRTVSGSFTVSGAP----PDGDHARVFPSPKVTDLAEFI 769
QY 765 GGSILNLTWAPGDDYDGHGTAHYIIRISTSLDLRDKFNESQVNTALIPKEANSEEV 824
Db 770 -GDIYHLTWAPGKVLNDRGHAHYIIRMSQHPDLQEDFNFNATLVNASSLIPKEAGSKET 828
QY 825 FLFKPENITFENCITDLFIATQAVDKVLDKSEISNIARVSLFIPQTPPETSPDETAPC 884
Db 829 FRKPFKFIANGIQYIALQADNEASLTSEVSNIA-----QAVKLTSLSDSISALG 880
QY 885 PNIH-INSTIPGIHIL 899
Db 881 DDISASMTWGLTVI 896
```

RESULT 9

```
Q9EQR4 ID Q9EQR4 PRELIMINARY; PRT; 902 AA.
AC Q9EQR4;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE ENDOTHELIAL CHLORIDE CHANNEL.
GN CLCA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC STRAIN=C57/BL6;
RA Elble R.C., Pauli B.;
RT "A murine endothelial homolog of Lu-ECAM-1 associated with lung
RT melanoma metastasis.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF115852; AAG47626.1; -
DR InterPro: IPR003961; FN.III.
DR InterPro: IPR002035; WFA.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS0234; WFA; 1.
SQ SEQUENCE 902 AA; 99855 MW; F2BA57DAA5D4AAAD CRC64;
```

Query Match 49.6%; Score 2358.5; DB 11; Length 902;
Best Local Similarity 52.6%; Pred. No. 4.2e-160;
Matches 480; Conservative 146; Mismatches 256; Indels 31; Gaps 14;
QY 1 MGPFKSSVFLIHLLEGALSLSLIQLNNGNYEGIVVADPNVPEDETLLQIKDMWTQA 60

```
Db 1 MVPGLQVLELTLHLQONT--ESSMVHLNSNGYEGVWIAINPSVPEDERLIPSIKEMVTA 59
QY 161 SLYLFEATGRKRFYKNVAILIPETWTKADYVRPKLETYKNADVLVAESTPPGNDPEYTE 120
Db 160 STYLFEASQGRVYFRNISILVPMWTKSKSEYLMPKRESYDKADVIVADPHLQHGDDPYTL 119
QY 161 OMNCCEKGRHILTPDFIAGKKLAEGYCPQGRAPVHWAHLRWGVDFEYNNDEKFIYS-N 179
Db 120 QYQCGDGRGYIHFTPNFLTDNLRIYGRGRVFEVHWAHLRWGVDFEYNNDRFYSRK 179
QY 180 GRIOAVRCSAGITGTNVVKKCGQSCVTKRCTFNKVTGLYEKGEFVLSQROTEKASIMF 239
Db 180 NTEATRCFSASITGKKVVHECQSGCVTRACRRDSKTRLYEPKCTFTFDKIQTACASIMF 239
QY 240 AQHVDSTVEFCTEQNHKEAPNKNCKNLRSWTWIEIRDSDEFKKTTPM--TTOPPNPTF 297
Db 240 MQLNSVVEFCTENHNAAEAPNLQNCNRRTWDVIKASADFQNSPPMRCTEAPPPPTF 299
QY 298 SILIOGIRIVCLVLDKSGSMATGNRLNRLOAQOGLFLLQTVELGVSQWGVTFDSAAHVQS 357
Db 300 SLLKSRVVCVLDKSGMDKEDRLIRMQAAELYLQIIVEKESMVLGVLTFDSAAHIQ 359
QY 358 ELIOINSQSDRDTLAKRLPAAAGSTGICSLRSFAFTVIRKK-YPTDGSSEIVLLTDGEDN 416
Db 360 YLIKITSSDYOKITANLPQOATGCTSIHGLQAGFOAITSSDQSTSGSEIVLLTDEGN 419
QY 417 TISGCNEVQSGAIHTHTVALGSAQAELSELSKMTGGLOQTYASDQVQNNGLIDAFGALS 476
Db 420 GISSCEAVSRSGAIHTHTIALGSAARELETLSDMTGGGLRYFANKV--SSLIDAFSRIS 477
QY 477 SGNCAVSRISOLSKGLTLNLSQWNCVIVDSTVGKDTPLFTITWTPPOILLWDPSC 536
Db 478 STSGVSQALQLESKAFNVRAGAWINSTVPDSTVGNDFVITWVQKPEIILQDPKG 537
QY 597 QK--QGFFVVDK--NTKMAVLIQPIGAKVGTWKYSLOAS--SQTLLTAVTSRASNATLPPI 592
Db 538 KKYITSDFOQDKLIRNSARLIQIPGTAETGTWTYSITGKSQLITMIVTTRASPTMEPLL 597
QY 593 VTSKTNKDTSKFSPPLVYVNIIRQASPIRLASVTALIESVNGKTVLELDNGAGADAT 652
Db 598 ATAHMSQSTAQYPSRMIVYARVSGQLPVLGAVNTALIEAHEGHQVTLLELDNGAGADTV 657
QY 653 KDGVSRYFTYDTHNGRYSVKVRALGGVNAARRVIPQOSGALYIPGWLENDEIQW 712
Db 658 KNDGIYTRYFTYHNGRYSKLVQVQARKNKARLS-LRQNKSLYIPGYVENGKIVLNP 716
QY 713 RPEINKDDVQHKQVCFSTRSSGGSFVASDVPNAPIDP-----LPPGQITDLKAEIHGS 767
Db 717 RPDVQBEAIEATVEDFNRTVSGSFTVSGAP----PDGDHARVFPSPKVTDLAEFI-GD 771
QY 768 LINLTWAPGDDYDGHGTAHYIIRISTSLDLRDKFNESQVNTALIPKEANSEEVFLF 827
Db 772 YIQLTWAPGKVLNDRGHAHYIIRVSHQPLGLQEDFNFNATLVNASSLIPKEAGSKETPK 831
QY 828 KPNITFENGTDLFIATQAVDKVLDKSEISNIARVSLFIPQTPPETSPDETAPC 887
Db 832 KPETFKIANDTQYIALQAYNEAGLTSEVSNIAQAVKF-----TSLEDSISALGADI 883
QY 898 H-INSTIPGIHIL 899
Db 884 SAIMTVMGLAVI 896
```

RESULT 10

```
Q9R070 ID Q9R070 PRELIMINARY; PRT; 902 AA.
AC Q9R070;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CA(2+)-SENSITIVE CHLORIDE CHANNEL 2.
GN CLCA2 OR CACC.
```

Tue Apr 2 09:40:07 2002

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HAMMARY;
PC MEDLINE=20012773; PubMed=10544033;
RX Lee D., Ha S., Kho Y., Kim J., Cho K., Baik M., Choi Y.;
RT "Induction of mouse Ca(2+)-sensitive chloride channel 2 gene during
RT involution of mammary gland";
RL Biochem. Biophys. Res. Commun. 264:933-937(1999).
DR EMBL: AF108501; AAF12731.1; -;
DR MGD: MGI:1931471; Cloc2.
DR InterPro: IPR002035; vWFA.
DR SMART: SM00327; vWA; 1.
DR PROSITE: PS50234; vWFA; 1.
DR SEQUENCE 902 AA; 99866 MW; 0FB746CF5FB3D07F CRC64;

Query Match 49.5%; Score 2353.5; DB 11; Length 902;
Best Local Similarity 52.5%; Pred. No. 9.5e-160;
Matches 479; Conservative 145; Mismatches 238; Indels 31; Gaps 14;

QY 1 MGPFKSVFLLILHLEGALSNLSLQNNNGYEGIVADPNVPEDETLIIQOIKDMVTOA 60
DB 1 MYPGLQVLLFTLHLQNT-ESSMHLNSNGYEGVVIAINPSVPEDERLIPIKEMVTOA 59
QY 61 SLYLFEATGKRFYFNKVAILLIPETWKTADYVPKLEYTKNADVLVASSTPPGNDPEYTE 120
DB 60 STYLFESAQGRVYFNISILYPMTWKSKSEYLMKRESYDKADVIVADPHLOHGDPPYL 119
QY 121 QMNGCEKGERIHLTPDFTAGKLAEYGPQGRAFYHEWAHLRWGVDFEYNNDEKPYLS-N 179
DB 120 QYQCGDRGOYIHFTPNFLTDLNRIYGRGRVFEHWAHLRWGVDFEYNDVPFYISRK 179
QY 180 GRIQAVRCAGITGNVNVKKCGGSCYTKRCFVNKVTGLYKGCFFVLQSRQTEKASIMF 239
DB 180 NTIEATRCASITGKVKVHECGQSCVTRACRRDSKTRLYEPKCTTIPDKIOTAGASIMF 239
QY 240 AOHVDSIVFECTEONHNKEAPNKQKCNLRSTWIEIRDSPEKTKTPM--TQPPNPPT 297
DB 240 MQLNSVVEFCTENNHNAAEAPNLQKNCRRSTWDVIKASADFQNSPPMRGTEAPPPPTF 299
QY 298 SLIQIGQIRVCLVLDKSGSMATGNRLNRNLNQACQLFLLQTLVELGSGVWMTFDSAAHVQS 357
DB 300 SLLKSRVVVCLVLDKSGSMDKEDRIIRNQAELYLTQIVKESWGLVTFDFAAHIQN 359
QY 358 ELTIQNSGSDRDLAKRLPAASGGTSGCLRSATFVIRKK-YPTDGSSEIVLLTDGEDN 416
DB 360 YLIKITYSSDYQKITANLPQATGGTSGICHLQAGQATITSSDQSTSGSEIVLLTDGEDN 419
QY 417 TISGCFNEVRKSGAIHTHVALGSAQAQELLESKMTGGLOTYASDQVQNNGLIDAFGLS 476
DB 420 GISSCFEAVSRSGAIHTHIALGPSAARELETLSDMTGGLRFYANKHV--SSLDADFSSIS 477
QY 477 SGNAGVRSQIQIESKGLTILNSQWNGTVIVDSVTKGKDTFLITWTQPQILLMDPSG 536
DB 478 STSGVSQQALQLESKAFNVRAGAINSTVLVDSVTGNDTFVITWTVQKEPILQDPKG 537
QY 537 QK--QGCFVVDK-NTKMAYLIQPICAKVGWTKVYSLQAS-SQTLPLVTVSASNATLPPI 592
DB 538 KKYITSDFQDELNIRSARLIQPGATGTGTWYTSITGKSQILMTVTTRASPTMEPLL 597
QY 593 VTSKTNKDTSKFPSPVVIYNIROGASPILRASVTALIESVNGKTVLELLDNCAGADAT 652
DB 598 ATAHMSOSTAQPSRMVIVYRVSGFLPVLGADVATIIAEAHGHQVLTLELMDNCAGADTV 657
QY 653 KDCGVYSRYFTYDNGRYSVKVRALGCVNAARRRVIPQOSGALYIPGWIEENDIOWNP 712
DB 658 KNOGIYTYFTYDNGNGRYSLKVRQAKNKARLS-LRQKNKSLYIPGYVENGKIVLNPP 716
QY 713 RPEINKDDVQHKQVCFSTSGGSFVASDVPNAPIDP-----LPFGQITDLKAEIHGGS 767

DB	717	RPDVOEEAIEATVEDFNRTVSGSLTVSGAP----	PGDHARVFPKSVTDLEAEFI-GD	771
QY	768	LINLTWTAPGDDYDGHGTAHYIIRISTISILDRKFNESQVNTTALPIKEANSEEVFLF	827	
DB	772	YIQLTWTAPGKVLDKGRAHRYIIRVQSHPGLQEDFNNTVLNASSLIPKAGSKETFKF	831	
QY	828	KPENITFNGTDLFTAIQAVDKVLDLKSITSILARYSLFIPPTPPTSPDSETSAPCPNI	887	
DB	832	KPETKIANDTQLYTAIQAYNEAGLTSEVSNIAQVKE-----TSLEDSISALGADI	883	
QY	888	H-INSTIPGIHIL	899	
DB	884	SALSMVWGLAVI	896	

RESULT	11
088860	PRELIMINARY; PRT: 901 AA.
ID	088860;
AC	088860;
DT	01-NOV-1998 (TReMBLrel. 08, Created)
DT	01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DT	CHLORIDE CHANNEL CACC.
DE	CLCAL.
GN	Mus musculus (Mouse).
CS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC	NCBI_TaxID=10090;
CN	[1]
RX	SEQUENCE 901 AA, 100038 MW, 13DF068BF3337AAB CRC64;
RP	Reinhold L., Musante L., Cinti R., Moran O., Seri M., Galletta L.J.V.;
RA	"Characterization of a murine gene homologous to the bovine Cacc
RT	chloride channel.,"
RL	Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR	ENBL; AF052746; AAC35003.1;
DR	MGD; MGI:1316732; Clcal.
DR	InterPro; IPR002035; VWFA.
DR	SMART; SM00327; VWA; 1.
DR	PROSITE; PS0234; VWFA; 1.
DR	SEQUENCE 901 AA, 100038 MW, 13DF068BF3337AAB CRC64;

QY	1	MGPKSSVFILILHLLGALSLSLIQNLNNNGYEGIVVAIDPNVPDETLTIQIQKDMVTOA	60
DB	1	MVPGLOVLLFTLHLQNT-ESSVHLNSNGYEGVVAIINPSVDERLIPSIKEMVTOA	59
QY	61	SLYLFEATGRFRYKPKNVAILIPETWKTKADYVRPKLETYKNADVLVAESTPPGNDPEYTE	120
DB	60	STYLFEASQGRVYFRNISILVPMTWKSKPYLMKRESYDKADIVADPHLQHGDDPYTL	119
QY	121	OMNGCGEGERIHLPDPIAGKCLAEGYPOGRAPVFEWAHLRWGVDFEYNNDEKFKYLS-N	179
DB	120	QYQCGDRGQYIHTFPNELLTDLNRIYGRGRVFEWAHLRWGVDFEYNNVDPFYMSRK	179
QY	180	GRIQAVRCSAGITGNVVKKCGGSCYTKRCTFNKVTGLYEKGFVLOSROTEKASIMF	239
DB	180	NTPIATRCSTRITGNVNVHNGERCNCVTRACRDSKTRLYEPKCTFIPDKIQTAGASIMF	239
QY	240	AQHVDSIVFECTEONHNKEAPNKQKCNLRSTWEVIRDSDEDFKTTFM--TTOPPNPTF	297
DB	240	MQLNLSVVECTEKNHNAEAPNLQKMCNRRSTWDVIKTSADFQNAFPMRGTEAPPPPTF	299
QY	298	SLIQIGQITVCLVLDKSGSMATGNRLNRLNAGQLFLLQTVELGCSWGMVTFDSAAHVOS	357
DB	300	SLKSRRRVCLVLDKSGSMKEDRLIRNMQAELYLTIQVEKESWGLVTFDSAAHIQ	359
QY	358	ELIQNSGSDRDTLAKRLPAAASGGTSCISGLRSFTVIRKK-YPTDGSSEIVLLTGDEN	416

Query Match	49.4%;	Score	2351;	DB 11;	Length	901;			
Best Local Similarity	53.7%;	Pred. No.	1.4e-159;						
Matches	471;	Conservative	143;	Mismatches	235;	Indels	28;	Gaps	13;

us-09-049-696-41.rspt

Tue Apr 2 09:40:07 2002

Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 DT EMBL; AF001263; AAB86531.1; -
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE LUTECAM-1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 RN NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RA Eble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
 RA Goodwin A., Pauli B.O.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF001262; AAB86530.1; -
 DR InterPro; IPR002035; WFA.
 DR PROSITE; PS0234; WFA; 1.
 DR SMART; SM00327; WFA; 1.
 SQ SEQUENCE 820 AA; 91464 MW; 4E01330AEF5A1494 CRC64;

Query Match 44.8%; Score 2131; DB 6; Length 820;
 Best Local Similarity 53.9%; Pred. No. 6.9e-144;
 Matches 426; Conservative 125; Mismatches 222; Indels 16; Gaps 11;

QY 8 VFILILHLLGALSLSLIOLNNNGYEGIVVAIDPNVPEDETLIOQIKDMVTOASLYLFEA 67
 DB 8 ILFLTLHLLPG-MKSSWNLNNGYDGLVIAINSPVPEDEKLIENIKEMVTEASTYLFA 66
 QY 68 TGRFYFNKVAAILPETWTKADYVRPKLETYKNADVLVAESTPPGNDPEYTMQNGCE 127
 DB 67 TKRRVFNVSILIPMTWKSSEYFIPKQESYDQADVIVANPYLKYGDPTLYQGRCE 126
 QY 128 KGERIHLPDFTAGKLAEYCPQGRAFVHWAHLRWGVFDEYNDEKLYLS-NGRIOAVR 186
 DB 127 KGKIHFTNPLNTHNFIYSGRGRVFWHWAHLRWGVFDEYNDEKLYLS-NGRIOAVR 186
 QY 187 CSAGITGTWV-KKCGGSCYTKRCTFNKVTGLYKGGFVLQSRQTEKASIMFAQHVD 245
 DB 187 CSTHITGINVFKKPGGSCITSLCRDSOTGLYEAKCTFLPKKQAKESIMFPLSHS 246
 QY 246 IVEFCTEQHNHKEAPNKNQCNLRSTWETIRDSDEKTKTPTMT--TOPNPFTSLLOIG 303
 DB 247 VTEFCTEKHTHTNAPLNQNMCKNGKSTWVIMNSVDFQNTSPTEMNPTHTFSLLSK 306
 QY 304 ORVCLVLDKSGSMATGNRLNRLNAGOLFLLQTVELGSGWGMVTFDSAHHVQSELIOIN 363
 DB 307 ORVCLVLDKSGMSAEDRLFQMQAELYLQVIEKSLVGMVTFDSVAEIQNHLTRIT 366
 QY 364 SGSDRDTLAKRLPAASGGTSCSGLRSAP-TVIRKKYPTDGESEIVLLTDGEDNTISCF 422
 DB 367 DDNVYQKITAKLPQVANGGTSICRGLKAGFOAIIHSDQSTSGSEIILLTDGEDNEINSCF 426
 QY 423 NEVKOSGAIHTVALGPSAAOELELSKMTGGLOTYASDQVONNGLIDAFGALSNGAV 482
 DB 427 EDVRSKGAIIHTVALGPSAAKELETLSNMTGGYRFFANKDI--TGLTNAFRISSRSGSI 484
 QY 483 SORSIOLESKGLTLQNSQWNGTVIVDSTVGKDTLFLITWTTPQPOILLWDPGSGO--KQ 540
 DB 485 TQQAIOLESKALKITGRKRVNGTVVPDSTVGNDFVFWTWTIQKPEIVLODPKGRKYTS 544
 QY 541 GFVVDK-NTKMAYLQIPGAKVGTWKYSL--QASSOTLTLTVTSTRASNATLPPITVTSK 596
 DB 545 DFKEDKLNIRSRARLOIPGIAETGTWYSLNNHASSOMLTVTTRARSPITPPVIATAH 604
 QY 597 TNKDTSKFPSPVVIYANIROGASPIILRASVTALIESVNGKVTLELLDNGAGADATKDDG 656
 DB 605 MSQHTAHYSPMIYVAYQSGFLPVLGISVIAIETEDGHQVTLLELDWNGAGRDVTYKNDG 664
 QY 657 VYSRYFTYDNGRYSVKVRALGVNAARRVIPPQSGALYIPGWENDEIONNPPREI 716
 DB 665 IYSRYFTDYNGRYSLKVAHAQARNNTARLNRPQNKVLYVPGVYENGKIILNPPREI 724
 QY 717 NKDDVOHKVCFSTSGGSFVASDV--PNAPIDLPFPQGITLQKAEIHGGSINLTWTA 775
 DB 725 KDLAKAKIEDFSRLTSGGSFTVSGAPPNGNHPSPVFPKSTIDLEAKFK-EDYIQLSWTA 783
 QY 776 PGDDYDHGTA 785
 DB 784 PGNVLDKGA 793

RESULT 14
 O18742 PRELIMINARY; PRT: 794 AA.
 ID O18742
 AC O18742;

DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE LUTECAM-1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 RN NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RA Eble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
 RA Goodwin A., Pauli B.O.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF001262; AAB86530.1; -
 DR InterPro; IPR002035; WFA.
 DR PROSITE; PS0234; WFA; 1.
 DR SMART; SM00327; WFA; 1.
 SQ SEQUENCE 794 AA; 88509 MW; B695E7256FC2C632 CRC64;

Query Match 44.7%; Score 2128; DB 6; Length 794;
 Best Local Similarity 53.9%; Pred. No. 1.1e-143;
 Matches 425; Conservative 125; Mismatches 222; Indels 16; Gaps 11;

QY 8 VFILILHLLGALSLSLIOLNNNGYEGIVVAIDPNVPEDETLIOQIKDMVTOASLYLFEA 67
 DB 8 ILFLTLHLLPG-MKSSWNLNNGYDGLVIAINSPVPEDEKLIENIKEMVTEASTYLFA 66
 QY 68 TGRFYFNKVAAILPETWTKADYVRPKLETYKNADVLVAESTPPGNDPEYTMQNGCE 127
 DB 67 TKRRVFNVSILIPMTWKSSEYFIPKQESYDQADVIVANPYLKYGDPTLYQGRCE 126
 QY 128 KGERIHLPDFTAGKLAEYCPQGRAFVHWAHLRWGVFDEYNDEKLYLS-NGRIOAVR 186
 DB 127 KGKIHFTNPLNTHNFIYSGRGRVFWHWAHLRWGVFDEYNDEKLYLS-NGRIOAVR 186
 QY 187 CSAGITGTWV-KKCGGSCYTKRCTFNKVTGLYKGGFVLQSRQTEKASIMFAQHVD 245
 DB 187 CSTHITGINVFKKPGGSCITSLCRDSOTGLYEAKCTFLPKKQAKESIMFPLSHS 246
 QY 246 IVEFCTEQHNHKEAPNKNQCNLRSTWETIRDSDEKTKTPTMT--TOPNPFTSLLOIG 303
 DB 247 VTEFCTEKHTHTNAPLNQNMCKNGKSTWVIMNSVDFQNTSPTEMNPTHTFSLLSK 306
 QY 304 ORVCLVLDKSGSMATGNRLNRLNAGOLFLLQTVELGSGWGMVTFDSAHHVQSELIOIN 363
 DB 307 ORVCLVLDKSGMSAEDRLFQMQAELYLQVIEKSLVGMVTFDSVAEIQNHLTRIT 366
 QY 364 SGSDRDTLAKRLPAASGGTSCSGLRSAP-TVIRKKYPTDGESEIVLLTDGEDNTISCF 422
 DB 367 DDNVYQKITAKLPQVANGGTSICRGLKAGFOAIIHSDQSTSGSEIILLTDGEDNEINSCF 426
 QY 423 NEVKOSGAIHTVALGPSAAOELELSKMTGGLOTYASDQVONNGLIDAFGALSNGAV 482
 DB 427 EDVRSKGAIIHTVALGPSAAKELETLSNMTGGYRFFANKDI--TGLTNAFRISSRSGSI 484
 QY 483 SORSIOLESKGLTLQNSQWNGTVIVDSTVGKDTLFLITWTTPQPOILLWDPGSGO--KQ 540
 DB 485 TQQAIOLESKALKITGRKRVNGTVVPDSTVGNDFVFWTWTIQKPEIVLODPKGRKYTS 544
 QY 541 GFVVDK-NTKMAYLQIPGAKVGTWKYSL--QASSOTLTLTVTSTRASNATLPPITVTSK 596
 DB 545 DFKEDKLNIRSRARLOIPGIAETGTWYSLNNHASSOMLTVTTRARSPITPPVIATAH 604
 QY 597 TNKDTSKFPSPVVIYANIROGASPIILRASVTALIESVNGKVTLELLDNGAGADATKDDG 656
 DB 605 MSQHTAHYSPMIYVAYQSGFLPVLGISVIAIETEDGHQVTLLELDWNGAGRDVTYKNDG 664
 QY 657 VYSRYFTYDNGRYSVKVRALGVNAARRVIPPQSGALYIPGWENDEIONNPPREI 716
 DB 665 IYSRYFTDYNGRYSLKVAHAQARNNTARLNRPQNKVLYVPGVYENGKIILNPPREI 724

us-09-049-696-41.rsp

Tue Apr 2 09:40:07 2002

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 04:17:57 ; Search time 69.35 seconds
(without alignments)
483.225 Million cell updates/sec

Title: US-09-049-696-41
Perfect score: 4759
Sequence: 1 MGPKSSVFILLHLEGAL.....GIHLKIMKWIGELQLSIA 914

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2411.5	50.7	903	1 ECLC_BOVIN	P54281 bos taurus
2	171.5	3.6	420	1 Y103_SYNY3	Q55874 synchocyst
3	153.5	3.2	1672	1 PMPB_CHLMU	Q9pjv2 chlamydia m
4	136.5	2.9	686	1 YPR4_CAEEL	Q20059 caenorhabdi
5	136	2.9	1849	1 IGA0_HAEIN	P45386 haemophilus
6	135.5	2.8	946	1 ITH2_MESAU	P97279-mesocricetu
7	133.5	2.8	1569	1 YPJA_ECOLI	P52143 escherichia
8	133	2.8	935	1 EAE_ECOLI	Q31000 escherichia
9	131	2.8	497	1 FLIC_ECOLI	P04949 escherichia
10	130.5	2.7	504	1 FLIC_SALMU	P06177 salmonella
11	130.5	2.7	976	1 PMPE_CHLMU	Q9pi47 chlamydia m
12	129.5	2.7	4590	1 FATH_HUMAN	Q14517 homo sapien
13	129	2.7	1493	1 NEO1_MOUSE	P97798 mus musculu
14	128.5	2.7	1754	1 PMPB_CHLTR	O84418 chlamydia t
15	128	2.7	493	1 FLIC_SALPA	P06178 salmonella
16	127.5	2.7	492	1 FLIC_SALRU	P06175 salmonella
17	127.5	2.7	1140	1 YM96_YEAST	Q04893 saccharomyc
18	127.5	2.7	1460	1 PMPC_CHLMU	Q9pjy1 chlamydia m
19	126.5	2.7	575	1 YFBK_ECOLI	P76481 escherichia
20	126.5	2.7	627	1 FLGK_BORBU	P70859 borrelia bu
21	126.5	2.7	1702	1 IGA2_HAEIN	P45384 haemophilus
22	126	2.6	489	1 FLIC_SALTY	P06179 salmonella
23	125.5	2.6	500	1 FLJB_SALAE	P52615 salmonella
24	125.5	2.6	1887	1 RPBI_DROME	P04052 drosophila
25	124	2.6	986	1 GUNZ_CLOSR	P23659 clostridium
26	124	2.6	1912	1 MTPD_HUMAN	P23468 homo sapien
27	122.5	2.6	1218	1 MGPC_MYCPN	O50341 mycoplasma
28	121.5	2.6	2944	1 CA17_HUMAN	Q02388 homo sapien
29	121	2.5	710	1 FIBI_ADEGI	O64761 avian adeno
30	121	2.5	1246	1 VP03_HSVSA	Q01000 herpesvirus
31	121	2.5	1541	1 IGA1_HAEIN	P42782 haemophilus
32	120.5	2.5	417	1 MAC1_YEAST	P35192 saccharomyc
33	120.5	2.5	591	1 FLGE_CAUCR	P35806 caulobacter

ALIGNMENTS

RESULT 1

ID	ECLC_BOVIN	STANDARD;	PRT;	903 AA.
AC	P54281;	1996 (Rel. 34, Created)		
DT	01-OCT-1996	(Rel. 34, Last sequence update)		
DT	01-OCT-1996	(Rel. 34, Last annotation update)		
DT	20-AUG-2001	(Rel. 40, Last annotation update)		
DE	EPITHELIAL CHLORIDE CHANNEL PROTEIN (CALCIUM-ACTIVATED CHLORIDE CHANNEL).			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.			
OC	NCBI_TaxID=9913;			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Trachea;			
RX	MEDLINE=96125078; PubMed=8537359;			
RA	Cunningham S.A., Awayda M.S., Bubien J.K., Ismailov I.I., Artate M.P., Berdiev B.K., Benos D.J., Fuller C.M.;			
PA	"Cloning of an epithelial chloride channel from bovine trachea.";			
FT	J. Biol. Chem. 270:31016-31026(1995).			
RL	FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME; MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND TRANSEPITHELIAL TRANSPORT.			
CC	SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).			
CC	TISSUE SPECIFICITY: TRACHEA.			
CC	PTM: PHOSPHORYLATED BY CAM-KINASE II.			
CC	SIMILARITY: CONTAINS 1 VWFA DOMAIN.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL: U36445; AAC48511.1; InterPro: IPR002035; VWFA.			
DR	SMART: SM00327; VWFA: 1.			
DR	PROSITE: PS50234; VWFA: 1.			
KW	Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;			
KW	Calcium channel; Phosphorylation; Glycoprotein.			
FT	TRANSMEM	7	27	POTENTIAL.
FT	TRANSMEM	331	351	POTENTIAL.
FT	TRANSMEM	617	637	POTENTIAL.
FT	TRANSMEM	883	903	POTENTIAL.
FT	DOMAIN	308	476	VWFA.
FT	CARBOHYD	75	75	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	278	278	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	360	360	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	372	372	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	504	504	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	515	515	N-LINKED (GLCNAC. . .) (POTENTIAL).

34	120	2.5	505	1 FLJB_SALTY	P52616 salmonella
35	120	2.5	928	1 HXA2_HAEIN	P45354 haemophilus
36	119.5	2.5	532	1 SPG7_DICDI	P22698 dictyosteli
37	119.5	2.5	1608	1 HLYA_SERMA	P15320 serratia ma
38	119	2.5	881	1 YJH8_YEAST	P47033 saccharomyc
39	119	2.5	1577	1 HLYA_PROMI	P16466 proteus mir
40	119	2.5	1609	1 PM21_CHLPN	O92605 chlamydia p
41	119	2.5	2334	1 WAPA_BAGSU	Q07833 bacillus su
42	118.5	2.5	977	1 AZA1_MOUSE	P17426 mus musculu
43	118.5	2.5	1276	1 PMP6_CHLPN	Q92899 chlamydia p
44	118.5	2.5	1694	1 IGA0_HAEIN	P44969 haemophilus
45	118	2.5	934	1 EAE_ECO57	P43261 escherichia

```
FT CARBOHYD 688 688 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 811 811 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 816 816 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 842 842 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 857 857 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 903 AA; 100305 MW; 9742AB1590908ABC CRC64;

Query Match 50.7%; Score 2411.5; DB 1; Length 903;
Best Local Similarity 54.2%; Pred. No. 3.9e-146;
Matches 483; Conservative 144; Mismatches 239; Indels 25; Gaps 13;

QY 1 MGPFKSVFLTLHLLEGALSNIOLNNNGYEGIVAIIDPNVPEDETLIOIKDMVTOA 60
Db 1 MVRPLVILFLTLHLPG-MKSMVNLINNGYDGIIVAIINPSVPEDEKLIQIKEMVTEA 59
QY 61 SLVLFATGKRFYKFNVAIIPETWTKTADYVRPKLETYKNADVLVAESTPNDPEYTE 120
Db 60 STYLFATKRRVVRNVSILPMTWKSSEYLMFKQESYDQAEVIVANPYLKHGDDPYTL 119
QY 121 OMGNCGEKGERHLTPDFTAGKLAIEYDPOGRAFYVHEWAHLRWGVFDEYNDEKPYLS-N 179
Db 120 QYGRCEKGGYTHFTPNFLTLNLPYIGSRGAFVHEWAHLRWGTFDEYNGDQPFYISR 179
QY 180 GRIQVRCAGITGNVVKCKGGSCYTKRCTFNKVTGLYKGCFFVLQSRQTEKASIMF 239
Db 180 NTIATRCSTHITGNVIVKCGGSCITRPCRRDSOTGLYEAKTFFPEKSQTARESIMF 239
QY 240 AQHVDIVSEFTEQHNKEAPKQKONLSTWIEVIRDSDFKKTTPMT--TQPPNPTF 297
Db 240 MGLSHVTEFTEKTHYENAPLNQKMGKSTWDVIMNSTDFQNTSPMTENPPTQPTF 299
QY 298 SLIQIGQRTVCLVLDKSGMATGNRLNLAQGLFLLOTVELGSGVGMVTFDSAAHVQS 357
Db 300 SLLKSKQVCLVLDKSGMSEDRFLFRNOAAELFQIEKSLGSLVGMVTFDSVAEIRN 359
QY 358 ELIQINSGSDRTTLAKRLPAAAGSTICSGLSAF-TVIRKKYPTDGESEIVLLTDCEDN 416
Db 360 NLTKITDDNVYENITANLPQANGSTICRGLKAGFQAIQSQOSTSGSEIILLTDCEDN 419
QY 417 TISGCFNEVKOGGALIHVALGSAAGLELKSMTGLQGLQYASDQVONNGLIDAFGALS 476
Db 420 EIHSCIEBKSGVYIHTIALGPSNAKELEFLTSDMTGHRFYANKDIF--NGLTNAFSRIS 477
QY 477 SGNCAVSORSIQLESKGLTLONSOMNGTVIVDSTVGKDTFLITWTQPPQILLWDPSG 536
Db 478 SRSGSITQOTIQLESKALATEKKWNGTVVDSTIGNDFTFVVTWTKPEILLQDPKG 537
QY 537 Q--KQGGFVVDK-NTKAYLIQIPGIAKVGTKYSL---QASSQTLTIVTSRASNTLPP 590
Db 538 KKYKTSDFKEDKLNHSARLRIPGIAETGWTYSLNNHASPOQLIATVITTRARSPTTP 597
QY 591 ITVTSKTNKDTKFPPLVYVYANIROGASPILRASVTALIESVNGKTVTLLELLDNGAG 650
Db 598 VTATAHMSQNTAHYSPSVIVYAOVSQGLFVLGINVTALIEEDGHQVLELDNGAGAD 657
QY 651 ATQDGVYRYFTYDNGRYSKVYALGVNNAARRVTPQSGALYIPGWIENDEIOWN 710
Db 658 TVANDGIYSRYFTYDNGRYSLKHAEARNNTARLSRQPNKALYIPGYLENGKIILN 717
QY 711 PPRPEIKDDVOHQV-CFSRTSSGSGFVADVPNAPIDLPFPQGITDLKA---EIHGG 766
Db 718 PPRPEV-KDLAKAEIDFSRLTSGSFVSGAPPNGNHPSPVLPNNKIIDLEAKFEDH-- 774
QY 767 SLINLTWAPGDYDGHGTAHYIIRISTSLDRKFNESLQVNTALITPKEANSEEVFL 826
Db 775 --IQLSWTAPANVLDKGRANSYIIRISKFLDQKDFDNATLVNTSSLKPKAGSDENFE 832
QY 827 FKPNITFENGDLFTAIQAVDKVDLKSFRISNIAVSLFIPQTPPTPSP 877
Db 833 FKPEPRIENGTFYIAVQAINFANLITSEVSNIAIKFIP---MPEDSV 880
```

```
RESULT 2
Y103-SYNY3
ID Y103-SYNY3 STANDARD; PRT; 420 AA.
AC Q55874;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 45.8 KDA PROTEIN SLL0103.
GN SLL0103.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RP SEQUENCE FROM N.A.
RA MEGLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugiura M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64 to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -!- SIMILARITY: TO E.COLI YPBK.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
-----
CC EMBL; D64004; BAA10635.1;
CC DR InterPro: IPR002035; WFA.
CC DR Pfam: PF00092; WFA; 1.
CC DR SMART; SM00327; WFA; 1.
CC DR PROSITE; PS50234; WFA; 1.
CC DR Hypothetical protein; Complete proteome.
CC SEQUENCE 420 AA; 45849 MW; E711B51478E74F3 CRC64;
-----
```

```
Query Match 3.6%; Score 171.5; DB 1; Length 420;
Best Local Similarity 24.0%; Pred. No. 0.00093;
Matches 109; Conservative 67; Mismatches 164; Indels 115; Gaps 22;

QY 251 TEQHNKAPKQKNCMLRSTWIEVRDSEDFKKTTPMTQPPNPTSLLOIGORIVCLV 310
Db 10 SDONLDAGAPT SQQ---LRIA--VAAKADHRRRLPLN-----LCLV 47
QY 311 LDKSGSMATGNRLNLAQGLFLLOTVELGSGVGMVTFDSAAHVQSELIQINSGSDRT 370
Db 48 LDHSGSM-DGQPLETVKSA--LGLIDRLIEDDRLSVIAFOHRAKIVIEVQVRNGAAIAK 105
QY 371 LAKRLPAAAGSTICSGLSAFTVIRKKYPTDGESEIVLLTDCEDNTISGCFNEVKOSGA 430
Db 106 AIERL--KAEGTAIDELKLGIOEAAKGEDRVSHFLITDGENE--HGDNDRCILKGT 161
QY 431 I-----IHTVALGPSAAQE-LEELSKMTGGLQTYASD-----QVONNGLI 469
Db 162 VASDYKLTHTLGLGHWNDVLEAIAAASQSLSYTENPSEALHTFRQLFQMSNVGLT 221
QY 470 DAFGALSNGA-----VSORSIQLESKGLTLONSQWNGTVIVDSTVGKDTFLITW 522
Db 222 NAHLLLELAQALAIKVPVQ--VSPETMDLTVNQGAIEEVRVLGLDMDTQDREVLN- 278
QY 523 TTOPPQILLWDPSGQKGGFVVDKNTKMYIQ--IPCIKVG--TWKYSIQASSQT---- 574
Db 279 -----LYLDOLLEGHQVIGOVQIRYDDPAGSQTNLLS 310
QY 575 ----LTLTTSRASNTLPIITVTSTKNDTSKFPSPLV-----VYANTROGASPIIR-AS 625
Db 311 DPLPLIOTQOTPDSTDOVOESILT---LAKYRQTOIAETKLRAGDRQGAATMLQTA 367
QY 626 VTAL-IESVNGKTVTLLELLDNG-----AGADATKDD 655
```

Db	275	DSOEIVP	-----SYNKASKDGGAI-----FAEKNVSEFNITTLK	308
QY	463	VQNGLI	DAFGALSSGNGAVSQRSIQLESKGLTLQNSQMMNGTVIVDSTVGKDFLITW	522
Db	309	VQNN	-----GAEEKGGIYASGDLSTOSSQSFLNS-----	339
QY	523	TTOPPQ	ILLWDPSQKGGFVVDKNTWKMYLQIPGIAKVTWKYSLQASSOTFLVTYSR	582
Db	340	-----	NTSKQGGALYIEGNVDFKLEETIRI-----KYNKSGTFFETKVTLSLP	383
QY	583	ASNATLP	PIVTSKNTDKTSKFFSPLY-----VYANIROGASPILRASVTALIESVNGK	636
Db	384	EATNKS	SVTAASOSGPNITPTTPPVATAKGGGLYTEKNLIS-----NITGIIETNNK	438
QY	637	TVTLELD	NGAGA-----DATKDDGVYSRYFTYDTNGRYSVKVR	676
Db	439	AT	-----DVGGAIVKGLTCKDSHRLOFKNSEKGGGLY-----TEDI	483
QY	677	ALGVNNA	RRVPIQOOSGALYIPGWIEDEIQNPPRPEINKDD-----VOHKOVCF	728
Db	484	NLTGKT	LFQENTAKEEGGLYIQG-----DDKTLTMTGLDSFCLIDN	525
QY	729	-SRTSGG	SFVSDVPNAPIDL--FPPGOITDLKAEIHGGSNLINLTWAPGGDDYDHGTA	785
Db	526	TSATHG	GGAVTKEISYTSYSDVEEFP--GIT-----PVHGETIISGNKATGGSGGVCTK	579
QY	786	HKYIIR	STSIDLRDNFESLQVNTTALIPKEANSEEVFLFKPENITFFENGTDLFIAIQ	845
Db	500	HLVLSN	LQTSIS-----ENFASENGSG--ACT	605
QY	846	AVDKVDL	KSEISNIARVSLFIPPQTPPETPSPDTSAP-----CPNIHINSTI	893
Db	606	CPDNF	-----PAPTASTPSTNQTAAPKDDKDFLIDVYVSTTIDKNKAT	648
QY	894	---PGIHL	KMKWIGELQLS 912	
Db	649	KKGAGV	YAKKAKLSRIDELNIS 670	

RESULT : 4

XPRA_CAEEL	ID	YPR4_CAEEL	STANDARD;	PRT;	686 AA.
AC	Q20059;				
DT	15-JUL-1998	(Rel. 36, Created)			
DT	15-JUL-1998	(Rel. 36, Last sequence update)			
DT	15-JUL-1998	(Rel. 36, Last annotation update)			
DE	HYPOPHETICAL: 77.0 KDA TRP-ASP REPEATS CONTAINING PROTEIN F35G12.4 IN CHROMOSOME III.				
DE	F35G12.4.				
GN	Caenorhabditis elegans.				
CC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;				
CC	Rhabditidae; Peloderinae; Caenorhabditis.				
CC	NCBI_TaxID=6239;				
CC	113				
CC	SEQUENCE FROM N.A.				
CC	STRAIN=BRISTOL N2;				
CC	Chui C.;				
CC	Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.				
CC	!- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMSL; Z46242; CAA86335.1; --				
DR	WormPep; F35G12.4; CE03286.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF00400; WD40; 7.				
DR	PRINTS; PR00320; GPROTEINBRPT.				

```
A Comparative genetic study of serologically distinct Haemophilus influenzae type I immunoglobulin A1 proteases." ;
```

```
J. Bacteriol. 174:2913-2921(1992).
```

```
-!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A  
PRODUCING INTACT FC AND FAB FRAGMENTS.
```

```
-!- CATALYTIC ACTIVITY: CLEAVAGE OF IMMUNOGLOBULIN A MOLECULES AT  
CERTAIN PRO-I-XAA BONDS IN THE HINGE REGION. NO SMALL MOLECULE  
SUBSTRATES ARE KNOWN.
```

```
-!- SUBCELLULAR LOCATION: SECRETED.
```

```
-!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC  
SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE  
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
DOMAIN. "THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY  
SIMILARITY)."
```

```
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
```

```
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between The Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial/  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).
```

```
-----  
EMBL; M87491; AAA24968.1; --  
DR MEROPS; S06.001;  
InterPro; IPR000710; IgA_S6.  
Pfam; PF02395; IGAI; 1  
PRINTS; PR00921; IGASERPRTASE.  
Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.  
FT PROPE 1022 1849 HELPER PEPTIDE (POTENTIAL).  
FT ACT_SITE 299 299 PROBABLE.  
SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;
```

RESULT	5
IGA4_HAEIN	
ID	IGA4_HAEIN
AC	STANDARD;
DC	P45386;
DT	01-NOV-1995 (Rel. 32, Created)
DT	01-NOV-1995 (Rel. 32, Last sequence update)
DT	01-NOV-1995 (Rel. 32, Last annotation update)
DE	IMMUNOGLOBULIN A1 PROTEASE PRECURSOR (EC 3.4.21.72) (IGAL PROTEASE).
GN	IGA
OC	Haemophilus influenzae;
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC	Haemophilus.
OX	NCHI_taxid=727;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NHIT HK61;
RC	MEDLINE=92234949; PubMed=1373717;
RA	Poulsen K., Reinholdt J., Kilian M.

QY 488 QLESKGLTQNSOMNGTVIV--DSTVCKDTFLITWTPTOPPOILLWDPSCOKGGFV-- 543
 Db 681 NVTENGSDQNFULTGNTLNGDLNVEKGFLLGSRPTHARDIAGISSTKDKDPHETEN 740
 QY 544 -----VDKNTKMAIQTPIGAKVGTWKYISLQASSOTLTLTVTSRASN-----A 586
 Db 741 NEVVVEDDWINRNFKATMTMVTGNA-----SLYSGRNVANITSNITASNAQVHLYGK 793
 QY 587 TLPPIVTSK-----TKDYSKF-PSPLVYVYANIQGASPI-LRASVTALIES 632
 Db 794 TGDTVCVRSYDGYVTYCHNSNLSEKALNSFNPTNLGRNVNLNTASFTLGKANLFGYIQS 853
 QY 633 VNGKTVLE-----LLDNG-----ACADATKDDGVYRFTVTDNGRY 671
 Db 854 IGTSQVNLKNSHWLGTGNSVNVNOLNTNGHIHLNAQDANK-----VTTYNT----- 901
 QY 672 SVKVRALGVNAARRRVIPOOSGALYIPGWIENDEIQMNPFRPEINKDDVQHKQVCFST 731
 Db 902 -LTVNSLSG-----NGSFYV--WVDFTNKSN--KVVNK----- 931
 QY 732 SSGSFVASDVNPAPIDLPFGQITDLKAEIHGSLI-----NLWTAPGDDYD 781
 Db 932 SATGNFTL-----QVADKTGEPNHINELTFDASNATRNLEVTLANGSVD 976
 QY 782 HGTAHKYIIRISTSIDLR-----DKENESLQVNTTALIPKEANSEVFLKPKENITFENG 837
 Db 977 RG-AWKYKLRNVGRYDLYNEVEKRNQOT--VDPT-----NITTPND 1015
 QY 838 TDLFIAIQADVQKVLKSEISNIARVSIFIPQTP-----PEPSPDTSAP 883
 Db 1016 -----IOA-DAPSAQSNNEIARVETPPPPAPATASEQPT-RAETAQP 1063

RESULT 6
 ITH2_MESAU STANDARD; PRT: 946 AA.
 AC P97279;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE INTER-ALPHA-TRYPsin INHIBITOR HEAVY CHAIN H2 PRECURSOR (ITI HEAVY
 DE CHAIN H2) (HC2).
 GN ITH2.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=97420688; PubMed=9276673;
 RA Nakatani T., Suzuki Y., Yamamoto T., Sinohara H.;
 RT "Molecular cloning and sequencing of cDNAs encoding three heavy-chain
 RT precursors of the inter-alpha-trypsin inhibitor in Syrian hamster:
 RT implications for the evolution of the inter-alpha-trypsin inhibitor
 RT heavy chain family.";
 RL J. Biochem. 122:71-82(1997).
 CC -1- FUNCTION: MAY ACT AS A CARRIER OF HYALURONAN IN SERUM OR AS A
 CC BINDING PROTEIN BETWEEN HYALURONAN AND OTHER MATRIX PROTEIN,
 CC INCLUDING THOSE ON CELL SURFACES IN TISSUES TO REGULATE THE
 CC LOCALIZATION, SYNTHESIS AND DEGRADATION OF HYALURONAN WHICH ARE
 CC ESSENTIAL TO CELLS UNDERGOING BIOLOGICAL PROCESSES (BY
 CC SIMILARITY).
 CC -1- SUBUNIT: I-ALPHA-I PLASMA PROTEASE INHIBITORS ARE ASSEMBLED FROM
 CC ONE OR TWO HEAVY CHAINS (H1, H2 OR H3) AND ONE LIGHT CHAIN,
 CC BIKUNIN. INTER-ALPHA-INHIBITOR (I-ALPHA-I) IS COMPOSED OF H1, H2
 CC AND BIKUNIN. INTER-ALPHA-LIKE INHIBITOR (I-ALPHA-LI) OF H2 AND
 CC BIKUNIN, AND PRE-ALPHA-INHIBITOR (P-ALPHA-I) OF H3 AND BIKUNIN.
 CC -1- PTM: HEAVY CHAINS ARE INTERLINKED WITH BIKUNIN VIA A CHONDROITIN
 CC 4-SULFATE BRIDGE TO THEIR C-TERMINAL ASPARTATE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ITH FAMILY.

CC SIMILARITY: CONTAINS 1 VWFA DOMAIN.
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
 CC or send an email to license@isb-sib.ch).
 DR EM3L; D89286; BAA13939.1; -
 DR InterPro: IPR002035; VWFA.
 DR Pfam: PF000092; vwa; 1
 DR SMART: SM00327; VWA; 1.
 DR PROSITE: PS0234; VWFA; 1.
 KW Serine protease inhibitor; Repeat; signal; Multigene family;
 KW Glycoprotein;
 FT SIGNAL 1 18 POTENTIAL.
 FT PROPEP 19 54 BY SIMILARITY.
 FT CHAIN 55 702 INTER-ALPHA-TRYPsin INHIBITOR HEAVY CHAIN
 FT H2.
 FT PROPEP 703 946 BY SIMILARITY.
 FT DOMAIN 308 468 VWFA.
 FT CARBOHYD 118 118 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 263 263 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 445 445 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC...) (POTENTIAL).
 FT BINDING 702 702 CHONDROITIN 4-SULFATE, CROSS-LINK SITE
 FT (BY SIMILARITY).
 SQ SEQUENCE 946 AA; 106580 MW; CA8BF565458E7B2E CRC64;

Query Match 2.8%; Score 135.5; DB 1; Length 946;
 Best local similarity 20.1%; Pred. No. 0.62;
 Matches 203; Conservative 124; Mismatches 334; Indels 347; Gaps 54;
 QY 20 LSNLSIQ--LNNNGYEGIVVAIDNPVPEDE-----TLIOIKDMVTOASLY-L 64
 Db 85 MATTIOSKLVNSPQSNVDFVQIPKGFISNFTVNGITSTIREKTVGRALYSQ 144
 QY 65 FEATGRFYKFNAILIPETWTKADYVRPKLEYKKNADVLVAESTPPG---NDEPTEQ 121
 Db 145 ARAGK-----TAGVRSRTLDENFTEV--NIPPGAKVQELHYQE 185
 QY 122 M--GNCKEGERIHLTPDFTAGKKLAEYGP-----QGRAFVHEWAHLRWGVDFYNDKX 174
 Db 186 MKWRKLGSEYHKIHLQPCRLA--KHLEVNVMVIVELQGMRFH-----VPDTF----- 230
 QY 175 FYLSNGRIQAVR-CSAGITGTNVVK--CQGGSCYTKRCTFNKVTGLYKGCFFVLQSR 230
 Db 231 -----EGHFQGVPIVSKQKSHVSFKFTVAQQRK--PNCYTTAVDG----- 271
 QY 231 QTEKASIMFAQHVDSIVFEFCTEQNHNEAPKNQKCNLRSTWEVIRDSSEFKTTPTMT 290
 Db 272 -----ELVMYDVNRE-----EKVGELEVFNGYFVHFF 299
 QY 291 OPPNPTFSLIQGRIVCLVDKSGSMATGNRLNRLNOAGOLF--LQTVELSGVWGMV 348
 Db 300 APEN-----LDPIPKNLFIVDVGSM--WGIMKQTVAMKTLIDDLRTEDQFS---VVD 350
 QY 349 FDSAAYH--QSLFIQNSGSDRTLAKRL--PAAASGCTSCSLGSAFTVIRK----- 399
 Db 351 FNNVTRWRNDLVSNATKTQITD--AKRIEIKIQPSGNTINEALLRAIFILNEASNLGML 408
 QY 400 YPTDGSIVLLTDGEDNTISGCFNEVQSGAIHTVALGPSAAQELBELSKMTGLQTYA 459
 Db 409 NPDSVSLIVLVSDG-DFTV---GELK-----LSK-----IKNV 438
 QY 460 SDQVQNNGLIDAFG-----ALSSNGAVSQRISQLESKGLTLQNSQWNGTVIV 508
 Db 439 KQNIQDNISLSFLGIGFDVVDYDFLKLRLSNENRGIAQRI--YGNRDTSSQLKFFNQ--V 493
 QY 509 DSTVGKDTLFLITWTPTOPPOILLWDPSCOKGGFVVDKNTKMAIQTPIGTAUKVTKYSL 568

Db 494 STPLRNQVF-----NYPQASVD-----VTQNFHNFVGGSEIIVAGKYDPSK 537
QY 569 QASSOTLTLTVTSRASNLPPIT-----VTSKTKNDTSKFPSPVLYVYANIRQ----- 616
Db 538 LAEVQSI-ITATSTNTLVLVLETLQSMODLEFLSKRHADPNFTKLLWAYLTINQLAER 596
QY 617 GASPIIRASVTALIESVNGKTV-----TLELDNGAGADATKDDGVYSRYFT 663
Db 597 SLAP-----TAATKRKTITKILQMSLDHIVTPLTAWIENEAGDERMLADSP----- 644
QY 664 TYDTNGRYSVKVRALGGVNAARRRVIPOQ-----SGALY-----IPGW----- 701
Db 645 -----PODHSCCSGALYYGTFKVASIPSWSAPSPTPVM 678
QY 702 -----TENDEIOWNPPRPRINKDDVOH-----KOVCFSRSTSSGGSFVASDVPNPAP 748
Db 679 AMLAVGANRLESTPPPHVIRVENDPHFIYLPKSKQKNCIFNIDSE----- 723
QY 749 DLFPFGQITDLKAEIHGGSLINLTWAPGDDYDHGTAKHYI-----IRISTS 795
Db 724 -----PGKILSLVSDPESILYN-QOLIGAKKAENGKLTFTYFGKLGFFYQKEDMKIELSTE 778
QY 796 ILDLRKDNESLQVNTTALIPKEA-----NSEEVFLP--KPNITFENGTDLFIAQVADKY 850
Db 779 NITL-----STTSLSFWSDTAHLGNQRVLISVKRGSVTLTLNKEFFSV-LLHHV 831
QY 851 DLKSEISNIARVSLFIPQTPETPSPDTSAPC-----PNHINSTIPG 895
Db 832 WKHHPV-NVDFELGIYLP--TNKFSFSAHGLGQFMKNPNIHFNERPG 877

RESULT 7

YPUA_ECOLI STANDARD; PRT; 1569 AA.
AC P52143; P76610; P77017, P77019;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 98.4 KDA PROTEIN IN ALPA-GABD INTERGENIC REGION (F949).
GN YPUA OR B2647.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features";
RL DNA Res. 4:91-113(1997).
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE (POTENTIAL).
CC -!- SIMILARITY: STRONG, TO BORDETTELLA PERTACTIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U36840; AAA79815.1; ALT_SEQ.
DR EMBL; AE000350; AAC75695.1; -.
DR EMBL; D90889; BAA16514.1; ALT_INIT.
DR EMBL; D90890; BAA16518.1; ALT_INIT.
DR EcoGene; EG13213; YPIA.
KW Hypothetical protein; Outer membrane; Complete proteome.
SQ SEQUENCE 1569 AA; 162774 MW; 0D006A9A32154596 CRC64;

Query Match : 2.8%; Score 133.5; DB 1; Length 1569;
Best Local Similarity 20.3%; Pred. No. 1.8;
Matches 159; Conservative 94; Mismatches 328; Indels 201; Gaps 34;

QY 164 GVFDEYNDEKFLYNGRIQAVRCSAGITGTNVVKKCOGGCYTKRCFTFNKVTGLYKGC 223
Db 645 GKMDEVGKDVGLVNSAGTQTIYASATSDKANI-----KGG-----KQTVYGLATEANIESG 696
QY 224 EFLVLOSROTEKASIMFAQHVDIVSEFTEQHNHKEAPNKQKCNLRSTWEVIRDSDFK 283
Db 697 EQIVDGGSTEKT-----HINGTQ--TVQNYGKAI-----NTDIV 729
QY 284 KTPMTQPPNPTFFSLQIQORIVCLVDKSGMATNRLNLRNLAQAGOLFLLQVLELGSW 343
Db 730 SGLQIQIMANGTAEGSIINGSOV-----NEGGLAEN-----SVLNDGG-----TLDVREKESA 778
QY 344 VGMVTFDSSAAHVQSELIQINSGRDRLAKRLPAAASGTSICSG-----LRSAFTVIRKK 399
Db 779 TGIQSSQCALVATTRATVCTGRADGVAFSTEQCAANNILLANGVLTVESDTSDDTKQ 838
QY 400 YPTDGEIV-----LLTGDENTTSGCFNEVKQSGAIHIV--ALGPSAAQLELELS 449
Db 839 VNMGRREIVKTKATATGTTLTGGE-QIVBGMANETTINDGGTQTVSANGEAIKTKINEGG 897
QY 450 KWTGGIQTVASDOVQNNGL-----IDAFGALSSNGAVS-----ORSIOLESKGLTLQ 497
Db 898 TLTVNDGKATDIVQNSGAALQTSANGIEISGTHQYTFISGNLATNMLLENGG----- 953
QY 498 NSQWMNGTVIVDSTVGKDTFLITWTTPQILLMDPSGQKGGFVVDKNTKMYLQIPG 557
Db 554 NLIVLAGTEARDSTVCK-----GGAMQNLGQDSATK---VNSGG 989
QY 558 IAKVGTWKYSLOASQTLTLTV---TSRASNATLPPITVTSKTKNDTSKFPSPVLYVYANI 614
Db 990 QYTLGRSKDEFQALAAEDLQVAGGTAIYVAGTLADASVSGATGSLMLTP----- 1040
QY 615 RQASPILRASVTALIESVNGKTVTLELDNGAGADATKDD-----GVYSRYFTTY 665
Db 1041 RDNVTPVKLEGAVRITDSA---TLTL-----GNGVDTTLADLTAASRGSVLNSNCSAG 1092
QY 666 DINGRYSVKVRAL---GGVNAARRRVIPOGSGALYIPGHIENDEIQOWNPREINKDDVOH 723
Db 1093 TSNCERYVNSLLNDGDDVYLSAQTAAPATNGIY---NTLTNNELSG-----SGNFYLH 1143
QY 724 KOVCFSR-----TSSGGSF--VASDVPNAPIDLPFPFGQITDLKAEIHGGSLINLTWT 774
Db 1144 TNVAGSRDQLVNNNATGNFKIFVQDTGCVSPQSD---DAMTLVKT---GGDASFTLG 1196
QY 775 AGDDYDHGTAKHYIIRISTSIILDRKFNESLQVNTTALIPKEANSEVFLKRPENTTF 834
Db 1197 NTGFGFVLDGT-YYEVLK-----SDGNSNNMLTNDV-----KPNPDPINPKPD----- 1238
QY 835 ENGTDLFIAQVADKVDLKSEISNIARVSLFIPQTPETPSPDTSAPCNIHINSTIP 894
Db 1239 -----PKPDPK-----PDNPKRPPDPTPTPTVPEKRTIPSTA 1272
QY 895 GI 896

```
Db 1273 AV 1274
RESULT 8
EAE_ECOLI1
ID EAE_ECOLI1 STANDARD; PRT; 935 AA.
AC O31000;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTIMIN (ATTACHING AND EFFACING PROTEIN) (EAE PROTEIN).
GN EAE OR EAP.
OS Escherichia coli O111:H-.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=168927;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98187918; PubMed=9529069;
RA Voss E., Paton A.W., Manning P.A., Paton J.C.;
RT "Molecular analysis of Shiga toxicogenic Escherichia coli O111:H-
RT proteins which react with sera from patients with hemolytic-uremic
RT syndrome.";
RL Infect. Immun. 66:1467-1472(1998).
CC -!- FUNCTION: NECESSARY FOR THE PRODUCTION OF ATTACHING AND EFFACING
CC LESIONS ON TISSUE CULTURE CELLS. BELIEVED TO MEDIATE ADHERENCE.
CC -!- SUBCELLULAR LOCATION: OUTER SURFACE.
CC -!- SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 LYSM REPEAT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF025311; AAC69247.1; -.
DR InterPro; IPR003344; Big_1.
DR InterPro; IPR003343; Big_2.
DR InterPro; IPR003535; Intimin.
DR InterPro; IPR002482; LysM.
DR Pfam; PF02369; Big_1; 2.
DR Pfam; PF02368; Big_2; 1.
DR Pfam; PF01476; LysM; 1.
DR PRINTS; PR01369; INTIMIN.
DR SMART; SM00257; LysM; 1.
DR Outer membrane; Virulence.
KW REPEAT 65 113
KW LYSM.
FT REPEAT 65 113
SQ SEQUENCE 935 AA; 101570 MW; 406E79CDC07DEB11 CRC64;

Query Match 2.8%; Score 133; DB 1; Length 935;
Best Local Similarity 21.0%; Pred. No. 0.88;
Matches 120; Conservative 94; Mismatches 240; Indels 118; Gaps 30;

Qy 358 ELIQINSQ-----SDRDLAKRLPAASGGTSICGLSAFTVIRKKYPTDGEIVLTD 412
Db 436 DLVQRNNIILEYKKQDILSNIPHNDINGTEHSTQKIQ---LIVKSKYGLD--RIV---- 486
Qy 413 GEDNTISCFNEVKOSGAIHTVALGPSAAQLELEL--SKMTGGLOTVASDOVQNGGLID 470
Db 487 WDDSLR-----SQGGQIQHS---GSQSAQDYQAILPAYVQGGSNLYK-----VT 528
Qy 471 AFGALSSGNGAVSQRSIQLESKGLTLQNSOMNGTVIVDSTVGK-----DTLFLITWTTQ 525
Db 529 ARAYDRNGN---SSNNVQLTI--TVLSNGVQVQVGVDFDTADKTSKADGTEAITYA- 582
Qy 526 PPQILLWDPSPGKGGFVVVDKNTKMKAYLIQIPGIARVGVFWKYSLOASSOTLTLTVTTSRASN 585
Db 583 -----TVKKNQ--VTQANVPVSFNVISGTATLGANSATTDANGKA---TVTLKSS- 627
```



```
Query Match      2.78; Score 130.5; DB 1; Length 504;
Best Local Similarity 21.98; Pred. No. 0.5;
Matches 100; Conservative 59; Mismatches 199; Indels 99; Gaps 20;

QY 316 SMATGNRLNRLNQAOLFLLOTVELGSWGVMTDSDAAHVOSLQINSQSGSDRDTLAKRL 375
DQ 10 SLTQNNLNKQSA-----LG-----TALERSLSSGLRINSKAD-DAAGQ-- 47
QY 376 PAAASGGTSGSGLRSAPTIRKKYPTDGSSEIVLLTGDENITISGCFNEVKQ----- 427
DQ 48 -AIANRFTANIKLQTA-----SRNANDGISIAQTTEGALNEINNLRVRELAVQSANG 101
QY 428 SGAIHTVALGPSAAQLEELSKMTG-----GLQTYASDQVQNNGLIDAFALSSGNGAV 482
DQ 102 TNSQSLDLSIAQETQRLNEIDRVSGOTQFNGVKVLA-----QDNLTQIQVANDGETIDI 157
QY 483 SORSQLESKGLTLQNSQ-----WNGTVIDSTV---GKDTLFLTLTWTTPPQILLWDPS 535
DQ 158 DIKELSSATLGLKLNQDQATVETAVTDKTYKNGDTFI-----TAQSNTDIQTAIG 212
QY 536 GQKGGFVVDKNTK--MAYLQIPGIAKVGWTKYSIQASSQTLTUTVTSRASNAITLP----- 589
DQ 213 GGATGVGTGADIKFQDQYLDVKGASAGVYKATYDETTKKVINDTIDKTPATAEATAI 272
QY 590 -----PITVTSKTNKTSKPPSPPLVVYANIRQASPIRLASVTALIESVNGKVTVL 640
DQ 273 RCTATITHNQIAEVTKEGVDTTTAAQLAA-AGV--TGADKDNKTSIVKLKSFEDKNGKVI-- 328
QY 641 ELLDNGAGADATKDDGVYSRYFTYD-----TNGRYSVKVRL--GGVNAARR 686
DQ 329 ---DGGYAVKMGDD-----FATYDEKQVOLLNNHVTDGAGVLQTVAVKFGGAN-GKS 379
QY 687 RVIPQSGALYIPGWIENDEIQWNPPEINKDDVOH 723
DQ 380 EVVTATVKGTVLASLDLKHNFRTGELKEVNTDKTEN 416

RESULT 11
PMPE_CHLMU STANDARD; PRT; 976 AA.
AC O9PL47;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE OUTER MEMBRANE PROTEIN PMPE PRECURSOR (POLYMORPHIC MEMBRANE
DE PROTEIN E).
GN PMPE OR TC0261.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Nigg9;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RL pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002293; AAF39130.1; .
DR TIGR; TC0261; .
DR IngerPro; IPR003357; OMP.
DR Pfam; PF02385; OMP; 1.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 18 POTENTIAL.
FT CHAIN 19 976 PROBABLE OUTER MEMBRANE PROTEIN PMPE.
SQ SEQUENCE 976 AA; 105964 MW; 0BEA56158E3CEA9 CRC64;
-----
Query Match      2.78; Score 130.5; DB 1; Length 976;
Best Local Similarity 17.78; Pred. No. 1.4;
Matches 147; Conservative 112; Mismatches 306; Indels 267; Gaps 37;

QY 72 FYFNVAIIIPETWKTRADYVR--PKLETYNADVLVAES-----TPGNDPEYTE 120
DQ 195 FCFRN---LTPESGVIGSPQNTPTTKIHNTIGPVLFFENNTCHRLWTOTOPENE----- 146
QY 121 QMNCGEKGERIHLPDPFIAGKLALEYGPOGRAFWHWAHLRWGVFD---EYNNDERFY 176
DQ 147 --GNKAREGGAIHAGDVVYSNNQNLV-----GFIKNFAYVQGGAISANTFAYKENKSF 198
QY 177 LSNRIQAVRCSAGITGNVVKKCGGSCYTKRCTFNKVTGLYKGCCEVQLQSROTERAS 236
DQ 199 L-----CLNNSC-----IQTKGGKGG 215
QY 237 IMEAQHVDSIVEFCTEQNHKEAPNKQKCN---LRSTWEVIRDSFD-----FKK 284
DQ 216 AIY-----VSTSCSFENNKKDLFIQNGCAGGAIFPTCSLIGNQGDIVFVSNHGPKN 269
QY 285 TPTMTPPNPTFSLQIQIRIVCLVLDKSGMATGNRLNQAOLFLLOTVELGSWV 344
DQ 270 VDNATNESGD-----GGAIKVTRLDITNNGSQIFFSDNISR----- 306
QY 345 GWTFDSAAHVOSLQINSQSGDRDTLAKRLPAAASGGTSGSGLRSAPTIRKKYPTDG 404
DQ 307 ---NFGGAIHAPC-LHLVGNQ---PTYFTNNIANHTGGAIIYTGTFETS-----KISADH 353
QY 405 SEIVLLTDGEDNTISGCFNEVKQSGAIHT-----VALGSAQAQLEEL 448
DQ 354 HALIF-----DNISA--NATNADGSSNTNPPHNRNATMDNSAGGIELGAGKSG----- 401
QY 449 SKMTGGLOTYASDQVQNNGLIDAFALSSGNGAVSQRSLQLESKGLTLQNSQWNGTVIV 508
DQ 402 -----NLIFYDPIQVTNAGVTVDENKDSQTCVV-----ESGATVL 438
QY 509 DSTVGKDTLFLITWTTPQPIILLWDPSQKGGFVVDKNTKMA---YLOIPGIKVG--- 562
DQ 439 SADISQANL---QTKTPATLTL-----SHGLLCIEDRAQLTVNNFTQTGTVALGNGA 488
QY 563 ---TWKYSIQASSQTLTLTTSRAS-----NATLPPITVTSKTNKDTSKPSPPLVYVAN 613
DQ 489 VLSQHSHTTDTATQTPPTTTTDDASVTLNHLGNLPSIL-----KDGAEPLLVPEVIS 542
QY 614 IRQGASPIRLASVTALIESVNGKVTLELLDNGAGADATKDDGVYSRYFTYDYNCRYSV 673
DQ 543 TTQG-NTTYTSDTAASFSLNG--ATLSLID-----EDG-NSPENTDLSRALYQ 589
QY 674 KVRALGGVNAARRVIPPQSGALYIP-----GWIENDEIQWNPPE--PEINKDDV 721
DQ 530 PMLAISEASDNQLOQSESMDFSKVNVPHYWGQGLMTGWAKTENPTTTPATITDPKKAQ 649
QY 732 QHKQVCFESRTSGGSFVASDVPNAPI-----PDLFPFGQITDLKAE-----IHG 765
DQ 650 FHRLLTLTWLPAQ--YIPSPKHKSPLIANTLWNLNTLGNLTKNSSGQELLDRPFWGITG 707
QY 766 GSLINLTWTAQDDYD-----HCTAHKIIRISTSLDLDRKPF 803
DQ 708 GGLGMVYQEPKRPKHDPGHMHTSGYSAGMITGNHTFSLRFSQSYTKLNERY 759
```


Db 2706 -SEFYFTVSEDPV- -GTEDIRAE- - - - - HSGTVLSLVKNTFESNRDESFVDRQ 2758
Qy 122 MNCCEKGERIHLPDFIAGKLAIEYQPGRAFVHEMAHLRWGVDFEYNN- - - - -E 173
Db 2759 SGRLEKSLDHTTKWQFSILARCTQDDHEWAS- -VDVSIQVKDANDNSPVFESSPYE 2817
Qy 174 KFYLSN- - - - -GRIOAVRCSAGITGVNVKKCGGSCYTKRTFNKVTGLYKCEGFVLOS 229
Db 2818 AFIVENLPGGSRVTOIRASDADSTN- - - - -GOVMYS- - - - -LDQ 2852
Qy 230 ROTKASIMFAQHYD- - - - -SIVEFCTEONHN- - - - -KEAPNKONOKNLRSTWEV- - - - - 275
Db 2853 SOSVEVIESFAINMETGHTITLKELDEKRDNYOIKVVASDHGEKIQLSSTAIVDVTVTD 2912
Qy 276 IRDS- - - - -EDFKKTTPTMTOPPNTFSLQIGORIVCLVLDKSGSMATGNRLNRL- - - - - 326
Db 2913 VNSPPRFTABIKYKGTVSEDDPGGVITAIL- - - - -STTDADSEINRQVTFYF 2959
Qy 327 - - - - -NQAGOLELLOTVELGSHVGV- - - - - - - - - - -TFDSNAHVQSELIO 361
Db 2960 ITGDPDLQ- - - - -PAVETIQ- - - - -NEKVVYVKKPLDKREKRDNYLLTITADGTFFSKAIVEVRVLD 3017
Qy 362 INSGSD- - - - -RDTL- - - - -AKRLPAAASGCTSCSLRSFTV- - - - - - - - - - - 395
Db 3018 ANDNSPVEKTLYSDTIPEDVLPKLIQISATDADIRSNAEITYTLGSGAEKFLNPD 3077
Qy 396 - - - - -IRKYPDTGSE- - - - -IVLLTDGEDNTISCFNEVKOSGAIHITVALGPSAAQ- - - - - 443
Db 3078 TGLKSTPLDREOAVVHLLVRATDG- - - - -GGRF- - - - -COASIVVTLVEDVNDNAPEFSA 3129
Qy 444 - - - - -ELEELSKMGGLQYASDQVONNGLIDAFALSGSGNGAVSORS- - - - -TQLE- - - - - 490
Db 3130 DPYAITVFENTEPTGLTRVQDADAGLNKILYSLDSADGQFSINELSGIIOLEKPL 3189
Qy 491 - - - - -SKGLTLQNSQWANGTVIVD- - - - - - - - - - -STVGKDTLF 518
Db 3190 DRELQAVYTLKAVDQGLPRELT- - - - -ATGTVIVSVLDINDNPPVFEYREYCATVSEILV 3247
Qy 519 LITWTQPPQILLMDP- - - - - - - - - - -SQKOGGFFVVDKNTKMA- - - - - 551
Db 3248 - - - - -GTEVLQVYAAASRDIEANAEITYSIISGNEHGFSDSKTGAVFIIENLDYESSHEY 3303
Qy 552 YLOI- - - - -PGIAKVGTKWYSL- - - - -QASSQITLTVTSRASNATLPPITVTSKT 597
Db 3304 YLTVEATDGGPPLSDVATVNVNTDINDNTPVFSQDTYTTVISEDVLEQSVITVMA- - - - - 3361
Qy 598 NKDTSKPPSLVWYANI- - - - -ROGAS- - - - -PIL-RASVTALI- - - - -ESVNGKVTLELLONGA 647
Db 3362 - - - - -DDADGPSNHHIYSLIDNGSGSFIDPVRGEVKYTKLLDRETISGYLTIVQASDNGS 3419
Qy 648 G- - - - -ADATKDDGVYGRYFTTYDTNNGYSVYKVRALGGVN-AARRRVIPOOSG 694
Db 3420 PPRVNTTNTVINDVSDVNDNAPVFSR- - - - -GNYSVIIQENKPVGVSVLQVLVTVTDEDS 3471
Qy 695 ALYIPGRI- - - - -ENDE- - - - -IOWNPPEINKDDVQHKQ- - - - -VCFSTSSGGSFVSDVP 743
Db 3472 SHNGPPFFFTVTDKAEFVNPQGVLLTSSAIKREKDHLYLOQKVADNGKPOLSLT 3531
Qy 744 NAPI- - - - -PDLFPFGQITDLX- - - - -AEIHGSLINLTWAPGDVDDHGT- - - - - 784
Db 3532 YIDIRVIEESYPPA- - - - -ILPLEIFITSSGEEYSGGVIGKIHATDQDVDTLTYSLDPQMDN 3590
Qy 785 - - - - -AHK- - - - -YIIRIS- - - - -TSILD- - - - -RDKFNESLOWNTT 812
Db 3591 LFSVSTGGKLIHAKKLDIGQYLLNVSVTDGKFTTADITVHIRQVTOEMLNHTIAIRFA 3650
Qy 813 ALIPKEANSE- - - - -EYFLKPEKNITENGTDL- - - - -FIAIQAVDKVLDKSEISNARVSLFPPQ 869
Db 3651 NLTPEEFVGYWRNFORALNRILGVRNDIQIVSLQS- - - - -SEPHPLDVLVLFV- - - - - 3700
Qy 870 TPTPTSPDETSAPCPNIHNTSTP- - - - -GIHLKIM- - - - - - - - - - -WKWIGE 908

Db 3701 - - - - -EKPSAQISTKQLLHKINSSVTDIEEIGVIRILNVFQKLCAGLDCPWFQDE 3752
RESULT 113
NEOL_MOUSE
ID NEOL_MOUSE ; STANDARD; PRT; 1493 AA.
AC P97798;
DT 20 AUG-2001 (Rel. 40, Created)
DT 20 AUG-2001 (Rel. 40, Last sequence update)
DT 20 AUG-2001 (Rel. 40, Last annotation update)
DE NEOGENIN PRECURSOR.
GN NEOL OR NGN.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain;
RX MEDLINE=97407661; PubMed=92644410;
RA Keeling S.L.; Gad J.M.; Cooper H.M.;
RT "Mouse neogenin, a DCC-like molecule, has four splice variants and is
expressed widely in the adult mouse and during embryogenesis.";
RL Oncogene 15:691-700(1997).
CC FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS: 1 (SHOWN HERE), 2, 3, 4
AND 5; ARE PRODUCED BY ALTERNATIVE SPLICING. THE EXPRESSION OF
ISOFORMS 3, 4 AND 5 ARE DEVELOPMENTALLY REGULATED.
CC TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC DEVELOPMENTAL STAGE: EXPRESSED UBICITOUSLY THROUGHOUT THE MID TO
LATE STAGES OF GESTATION AND IN ADULT TISSUES. STRONG EXPRESSION
IS OBSERVED IN THE VENTRAL REGION OF THE VENTRICULAR ZONE OF THE
E15.5 MOUSE NEURAL TUBE, AS WELL AS IN THE VENTRICULAR ZONES OF
THE MESENCEPHALON AND RHOMBENCEPHALON. ISOFORMS 3 AND 4 ARE
EXPRESSED AT HIGHER LEVEL COMPARED TO OTHER ISOFORMS BETWEEN E11.5
AND E16.5.
CC SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. STRONG, TO
TUMOR SUPPRESSOR PROTEIN DCC.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
DR EMBL; Y09535; CAA70727.1; -
DR HSPB; P02751; ITTG
DR MGD; MGI-1097159; Neol.
DR InterPro; IPR001777; FN_III.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003598; Ig_C2.
DR Pfam; PF00041; fn3; 6.
DR Pfam; PF00047; ig; 4.
DR PRINTS; PR00014; FNYPEIII.
DR SMART; SM00060; FN3; 6.
DR SMART; SM00408; IGC2; 4.
DR Transmembrane; Immunoglobulin domain; Glycoprotein; Signal;
KW Alternative splicing.
FT SIGNAL 1 36
FT CHAIN 37 1493
FT DOMAIN 37 1136
FT TRANSMEM 1137 1157
FT DOMAIN 1158 1493
FT DOMAIN 178 147
FT DOMAIN 177 239
FT DOMAIN 177 239
FT DOMAIN 177 239

FT DOMAIN 274 338 IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN 366 428 IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN 467 564 FIBRONECTIN TYPE-III 1.
FT DOMAIN 567 660 FIBRONECTIN TYPE-III 2.
FT DOMAIN 661 760 FIBRONECTIN TYPE-III 3.
FT DOMAIN 766 860 FIBRONECTIN TYPE-III 4.
FT DOMAIN 881 981 FIBRONECTIN TYPE-III 5.
FT DOMAIN 982 1083 FIBRONECTIN TYPE-III 6.
FT DOMAIN 1149 1153 POLY-VAL.
FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 746 746 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 940 940 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 442 461 MISSING (IN ISOFORM 2).
FT VARSPLIC 863 878 MISSING (IN ISOFORM 3).
FT VARSPLIC 1086 1096 MISSING (IN ISOFORM 4).
FT VARSPLIC 1279 1331 MISSING (IN ISOFORM 5).
SQ SEQUENCE 1493 AA; 163159 MW; 441DE919D5E17C0E CRC64;

Query Match 2.7%; Score 129; DB 1; Length 1493;
Best Local Similarity 20.6%; Pred. No. 3.2;
Matches 153; Conservative 86; Mismatches 219; Indels 284; Gaps 42;

QY 292 PPNFTSIIOTGRIYGLVLDKSGMATNRLNRLNQAQOL-----FLQVVELGSWV 344
DB 21 PPPPLLLLLPL-----LLLGRPASGAATKSGPRQSGASVRTFTFFFLVEPV----- 71
QY 345 GMVFTDAAHVQSELIQNSGDRDTIAKRLPAAASGTSI-----CSGLKSAFTVIRKKYP 401
DB 72 -----DTLSVR-----GSSVILNCSAYSESPNI--EWK 98
QY 402 TDGSEIVLLTGDENTISGCFENEVKQSGA-----IIHTVALGSPMAQEELESKMTGGQL 456
DB 99 KDGFTFLNLSDDR-----ROLLPGSLFISNVHSHKHKDD-----EGFYQ 139
QY 457 TVASDVONNGLIDAFALSGSGGAVSORSIQLESKGLTLQNSO-----WMNGTVIVDS 510
DB 140 CVAT-----VDNLGTIVS-----RTAKLTAVGLPRFTSGPEPSSVYVGNLSAILNC 184
QY 511 TVGKDTLELIWTTOPPOILLWD-----PSG-----QKQGF-----VVDKNTKWAYLQ 554
DB 185 EYNADLVFPVRWEQNRQPLLLDDRIVLPLPSGTLVISNATEGDGLYRCIVESG----- 237
QY 555 IPGIKAVCTWKYS-----LQASSQTLTLTVTSRAS-----NATLPPITVTSKTK 599
DB 238 -----GPPFESDEAEKLVLDPEEIVDLVFLMRPSSMMKVTSQSAVLPCV----- 282
QY 600 DTSKPPSPLVYANIRQSGAPILRASVTALIESVNGTIVLELLDNAG-----ADATKD 654
DB 283 -VSLGPAPVVRMKNEE-----VLDTESSGRLVLL-----AGCLESIDVTE 324
QY 655 DGVISRYETTYDNGRYSVKVYKVALGGVNAARRVIP-----QQSGALY-----IP 699
DB 325 DA--GTIFCIAD--NGNKTVEQAELTV-----QVPPGLKOPANIYAHESMDIVFCEVT 376
QY 700 G-----WIENDEIOWPPRPETNKDVOHKQVCFSTSSGGSF-----VASDVNA----- 745
DB 377 GKPTPTVKVKNQGVVIFSDNFKIVKE--HNLQVLGLVKSDGEFYQCIAENDVGNQAAGA 434
QY 746 -----PTDLPFPQGIT-----DLKAEIHGGSLLNLTWAP 776
DB 435 QLIILEHDVAIPTL--PPTSLTSATTDHLAPATGPLSPAPRDVVASLVSRTFKLWRT 493
QY 777 GDYDHGTAKHYIIRISILDLRDKFNES-----LQVNTTALIPKEANSEEVFLKPENI 832
DB 494 ASD--PHGDNLTYSVFTYREGVDREVENTSQPGEMQVTLQNLMP-----ATVYIFK----- 543
QY 833 TFENGTDLIAQAVDKVDLAKSEISNIARVSLFIPPTQPTTPTSPDETSAPCPNIHINST 892

DB 544 -----VMAQNKHG--SGESSAPLRV-----ETOPEVQLP-----GPAPNIRAYAT 581
QY 893 IPGIHILKIMWK-----WIGELQ 910
DB 582 SP--TSITVITWEIPLSGNGEIQ 601
RESULT 14
PMPB_CHLTR
ID PMPB_CHLTR STANDARD; PRT; 1754 AA.
AC 084418; 2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE OUTER MEMBRANE PROTEIN PMPB PRECURSOR (POLYMORPHIC MEMBRANE PROTEIN B).
DE PMPB OR CT413.
GN Chlamydia trachomatis.
OS Chlamydia trachomatis.
CC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UV-3/CX;
EX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L., Kitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;
RA "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis";
RL Science 282:754-759(1998).
CC -! SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES) (POTENTIAL).
CC -! SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EM5L; AE001314; AAC68010.1; ALT_INIT.
DR InterPro: IPR003357; OMP.
DR InterPro: IPR003368; DUF145.
DR Pfam: PF02415; DUF145; 1.
DR Pfam: PF02385; OMP; 1.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 14 POTENTIAL.
FT CRAIN 15 1754 PROBABLE OUTER MEMBRANE PROTEIN PMPB.
SQ SEQUENCE 1754 AA; 183317 MW; 969CF8D8D36185D CRC64;
Query Match 2.7%; Score 128.5; DB 1; Length 1754;
Best Local Similarity 19.7%; Pred. No. 4.3;
Matches 161; Conservative 108; Mismatches 313; Indels 235; Gaps 38;
QY 211 TPKVITGLYEKCEVFVQSRQTEKASIMFAQHVDSIVFCTEQNHKAPKQKCNLR 270
DB 43 TFEITVG--EAGAEIYVSGNA-----SFTKFTNIPPTDITPTNSSSSSNG 87
QY 271 STWEVIRDSDEKFTTP-----MTQPPNPFTSLQI-----QQRIV 307
DB 438 ETASVSEDS--DSTTTTTPDPKGGGAFYNAHSGVLSEFMTSGTSCUTLSEIKITGEGGAIF 146
QY 308 ----CLVLQKSGMATGNRLNRLNQAQOLFLQTVELGSWVGWTFDS----- 351
DB 147 SQGELLFTDLTG--LITQNLNLSQLS--GGAIFGESTISL--SGITKATFSSNAEVPAPVKP 203
QY 352 ---AAHVQSELIQNSGSDRTL-----AKRLPAAASGCTSCSLGRLSFTVIRKKYPTD 403
DB 204 TEPKATASSETSGSSSGNDSSVSSPSSRAEPAAAN-----LQSHFICAT---ATP 252

Qy 404 GSEIVLLTDCEDNITSCFNEVKSGAIHTHTVALGPSAAOE-LEELSKMT--GG----- 454
 Db 253 AAQ-----TDTETSTPS---HRPGSGGAIYAKGDLTIADSOEVLFSINKATKDGGAFAEK 305
 Qy 455 ---LOTVASDOVQNNGLIDAFALSSGNGAVSORSIOLESKGLTLQNSOWM---NGTVIV 508
 Db 306 DVSFENITSLAVQIN-----GAEEKGAIYAKGDLSTOSSKQSLFNSYKSGGGALYV 359
 Qy 509 DSTVGKDTFLI-----TWTTPQQLLWDPGKQGGFVVDKNTKMYLQIPGIAKVG 562
 Db 360 EGDINFQDLEERIKYKAGTFETKKTIL--PKAASAG-----NAD 399
 Qy 563 TWKYSLOASQTLTIVTSRASNLTPITVTSKTNKDTSKFPSPVLYANIROGA----- 618
 Db 400 AWASSSPQSGGAT-TVNSNGDS-----SSGSDSDTSE-----TVPATAKGGGLYTD 445
 Qy 619 SPILRASVTALIESVNGR-----TWT-----LELLDNGAGADATKDDGVYSR 660
 Db 446 KNLSTNITGIIETANNKATDVGGGAYVKGTLTCTENSRLQFLKNSD---KQGGIY-- 500
 Qy 661 YFTTYDNGRYSVKVRALGVNAARRVIPPQSGALYIPGWIEDEIQWNPPEINKDD 720
 Db 501 -----GEDNITLSNLTKTLEQENTAKKEGGGLFKGTDKALTWTGLDSFCLINNTS 552
 Qy 721 VOHKQVCFRSTSGGSFVADVPNA-----PIPLPPGQITDLKAEIHGSLINLTWT 774
 Db 553 EKH-----GGGAFVKEISQTYSDVETIPGTP-----VHGETVITGNKS 593
 Qy 775 APGDDYDHGTAHKYIIRISTSLDLRDKFNESLQVNTTA-----LIPKEANSEE 823
 Db 594 TGGNGG-----VCTKRLALSNLQISISGNSAAENGGAHTCPDSFPTADTAEQ 643
 Qy 824 VLFKPNENITFENGTDLEIAQVDKVDLKEISNIARVSLFIPPQPPETPSPDETSAP 883
 Db 644 PAAASAASTPESAPVWSTALSTSSSTV-SLTLAASSQASPATSNKTKQDPNADTDL 702
 Qy 884 CPNTHINSTIP-----GHIHLKIMKWIGELQLS 912
 Db 703 LIDYVDTTISKNTAKKGGGIYAKKMSRIDQLNIS 739

RESULT 15
 ID FLIC_SALPA STANDARD; PRT: 493 AA.
 AC P06178;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE FLAGELLIN (PHASE-1-A FLAGELLIN).
 GN FLIC OR FLAF OR HAG.
 OS Salmonella paratyphi-a.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OC NCBI_TaxID=54388;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 9150;
 RX MEDLINE=86143833; PubMed=4093985;
 RA Wei L.-N., Joys T.M.;
 RT "Covalent structure of three phase-1 flagellar filament proteins of
 Salmonella";
 RL J. Mol. Biol. 186:791-803(1985).
 CC -|- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
 FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
 CC -|- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE
 BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLA, TERMED
 CC PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.
 CC -|- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its

used by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X03393; CAA27128.1; ALT_INIT.
 DR F18; S07276; S07276; Flagellin_C.
 DR InterPro; IPR001029; Flagellin_C.
 DR InterPro; IPR001492; Flagellin_N.
 DR Pfam; PF00700; Flagellin_C; 1.
 DR Pfam; PF00669; Flagellin_N; 1.
 DR PRINTS; PR00207; FLAGELLIN.
 DR PubMed; PM000316; Flagellin_C; 1.
 KW Flagella.
 ET INIT_MET : 0
 SQ SEQUENCE 493 AA; 51735 MW; CC657679B86C8374 CRC64;
 Query Match 2.78; Score 128; DB 1; Length 493;
 Best Local Similarity 22.88; Pred. No. 0.7;
 Matches 93; Conservative 58; Mismatches 165; Indels 92; Gaps 21;
 Qy 316 SMATGNRLNRLNQAGQLFLQTVELGSGWVGWVTFDSAAHVQSELIQINSGSDROTAKRL 375
 Db 310 SLLTQNNLKSQA-----LG-----TAIERLSSGLRLNSAKD-DAAGQ-- 47
 Qy 376 PAAASGTSICSGURSFAFTVIRKKYPTDGSSEIVLLTDCEDNITSCGFNEVKQ----- 427
 Db 448 -AIANRFETANIKGLTQA-----SRNANDGISIAQTTEGALNEINNNLQVRRELAVQSANS 101
 Qy 428 SCATHTVALGPSAAQLEELSKMTG-----GLQTYASDQVQNNGLIDAFALSSGNGAV 482
 Db 102 TNSQSDLDLSIAQETQRLNEIDRVSGQTQFNGVKVLA-----ODNTLTIQVANDGETIDI 157
 Qy 483 SQRSIQLESKGLTLQNSQ-----WMNGTVIVDSTVGKDTLFLITWTTPQQLLWD 533
 Db 458 DLKQINSQTLGLDITLNVQKKYDKVSEVTPSATL---STTALDAGLKTGTGCTT-----D 209
 Qy 554 PSGOKQGGFVVDKNTKMYLQI-----PGTAKVGTWKYSLOASQSLTIVTSRASNT 587
 Db 210 TGSIKDKGVYVNSTSKNYVEVEFTDATDQTNKGGFYKVVN-ADDGAVTMTAATTKKATT 268
 Qy 588 LPPITVTSKTNKDTSKFPSPVLYANIROGASPIILRASVTALIE-----SVNGKTVTLLELL 643
 Db 269 PTGITEVTQVOKPVA---APAAIQALF--AAHTGADTAEMVRKMSYTDKNGKTI----- 318
 Qy 644 DNG-----AGAD---ATKD-DGVYSRYETTY-DTNGRYSVKVRALGGVN 682
 Db 319 DGGFGVKGADIIYAATKNKDGSPSTINTTEYTDKDGNTKTALNQLGGAD 366
 Search completed: March 30, 2002, 04:24:11
 Job time: 374 sec

us-09-049-696-41.rsp

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 03:22:22 ; Search time 72.09 Seconds
(without alignments)
965.786 Million cell updates/sec

Title: US-09-049-696-41
Perfect score: 4759
Sequence: 1 MGPKSSVFILILHLLLEGAL.....GIHLKIMKWNIGELQLSIA 914

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3644.5	76.6	913	2 JG0168	gob-5 protein - mo
2	2334	49.0	905	2 T02205	Lu-ECAM-1 protein
3	171.5	3.6	420	2 S76691	hypothetical prote
4	153.5	3.2	1672	2 C81675	polymorphic membra
5	153	3.2	689	2 F84811	probable retroelem
6	152	3.2	2650	2 E85822	probable invasin 2
7	145.5	3.1	928	2 C81265	probable lipoprote
8	144.5	3.0	5188	2 B85547	probable RTX famil
9	143.5	3.0	2340	2 B71704	cell surface antig
10	138	2.9	683	2 T21810	hypothetical prote
11	136.5	2.9	686	2 T21808	hypothetical prote
12	136.5	2.9	1451	2 A85547	hypothetical prote
13	136	2.9	698	2 A69222	hypothetical prote
14	136	2.9	1849	2 C41859	IgA-specific metal
15	135.5	2.8	946	2 JC5575	inter-alpha-trypsi
16	135.5	2.8	978	2 D81411	probable lipoprote
17	135	2.8	1651	2 T41160	transmembrane rece
18	134.5	2.8	955	2 G64866	probable membrane
19	134	2.8	1173	2 I50620	procrar2 - chicken
20	134	2.8	26926	1 I38344	titin, cardiac mus
21	133.5	2.8	949	2 T08658	hypothetical prote
22	133.5	2.8	1569	2 A65044	hypothetical prote
23	133	2.8	1039	2 T09883	hypothetical prote
24	132	2.8	918	2 T02759	hypothetical prote
25	131.5	2.8	1742	2 S24600	projectin - fruit
26	131	2.8	498	1 FLEC	flagellin - Escher
27	131	2.8	1109	2 A56143	surface-array prot
28	130.5	2.7	505	2 S09638	flagellin - Salmon
29	130.5	2.7	976	2 F81722	polymorphic membra

30 130.5 2.7 4152 2 T31102 filamentous hemagg
31 130 2.7 1067 2 D75625 probable extracell
32 130 2.7 1441 2 B86807 hypothetical prote
33 130 2.7 1894 2 C54689 protein-tyrosine-p
34 130 2.7 2232 2 T34434 hypothetical prote
35 129.5 2.7 1104 2 S59310 probable membrane
36 129 2.7 1421 2 T02501 hypothetical prote
37 129 2.7 1528 2 D85912 hypothetical prote
38 128.5 2.7 1751 2 G71518 hypothetical prote
39 128 2.7 494 2 S07276 flagellin H-1A - S
40 128 2.7 676 2 T47637 flagellin H-1A - S
41 128 2.7 1474 2 F69009 hypothetical prote
42 128 2.7 1536 2 A43855 probable membrane
43 128 2.7 1612 2 T30805 high-molecular-wei
44 128 2.7 1691 1 D54689 outlin protein - mo
45 127.5 2.7 493 2 S07375 protein-tyrosine-p
flagellin H-1 - Sa

ALIGNMENTS

RESULT 1

JG0168

gob-5 protein - mouse

C:Species: Mus musculus (house mouse)

C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000

C:Accession: JG0168

R:Komiya, T.; Tanigawa, Y.; Hirohashi, S.

Biochem. Biophys. Res. Commun. 255, 347-351, 1999

A:Title: Cloning and identification of the gene gob-5, which is expressed in intestine

A:Reference number: JG0168; MUID:99160866

A:Accession: JG0168

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-913 <KOM>

A/Cross-references: DDBJ:AB016592

Query Match 76.6%; Score 3644.5; DB 2; Length 913;
Best Local Similarity 75.6%; Pred. No. 1.7e-232;
Matches 692; Conservative 102; Mismatches 114; Indels 7; Gaps 4;

QY	1	MGPKSSVFILILHLLLEGALNSLIQLNNGYEGIVVAIDNPVDEFLIQIKDMVTOA	60
DB	1	MESLKSVPVLLILHLLGVLSESLIQLNNGYEGIVVAIDHVPDEALIQIKDMVTOA	60
QY	61	SLYLFATGKRFYKFNVAAILIPETWKADYVRPKLETYKNADVLVAESTPPGNDEPYTE	120
DB	61	SPYLFATGKRFYKFNVAAILIPESWKAKPEYTRPKLETYKNADVLVSTPSLGNDEPYTE	120
QY	121	OMGNCGEGERIHLPDPIAGKLALEYGPQGRAFVHEWAHLRWGVDFEYNNDEKPYLSNG	180
DB	121	HIGACGEKGIHLPDPIAGKLTQYGPQDRTFVHEWAHFRWGVDFEYNNDEKPYLSNG	180
QY	181	RIQAVRSAGITGVNVKKCGGSCYTK-RCTFNKVTGLYKGCFFVLQSQTEKASIMF	239
DB	181	KQAVRCSAATGKNQVRRCGGSCITNGKVIDRTGLYKDNCFVFPDPHONKASIMF	240
QY	240	AQHVDSTVEFTEQNHKNEAPKNQKCNLRSTMEVIRDSDFKKTTPMTTPPPNPTFSL	299
DB	241	NONINSVVEFTEKHNHNOEAPNDQNCNLRSTMEVIOESDFKQTTPMTAQPAPTSL	300
QY	300	LQIGORIVCLVLDKSGSMATGNLRNLNQAGQLFLQTVLGSGWVGMVTFDAAHVQSEL	359
DB	301	LQIGORIVCLVLDKSGSMATGNLRNLNQAGQLFLQTVLGSGWVGMVTFDAAHVQSEL	360
QY	360	IQINSGSDRTLAKRLPAAASGGTSGSLRSATVIRKKYPTDGSSEIVLITDGEDNTIS	419
DB	361	KQLNSGADRLILKHLPTVTSAGGTSICSLRTAFVIRKKYPTDGSSEIVLITDGEDNTIS	420
QY	420	GCFNKGSGAIHTVALGPSAAOELSELSKMTGLQTYASDVQNNGLIDAFALSSGN	479
DB	421	SCFDLVKOSGAIHTVALGPSAAOELSELSKMTGLQTYASDVQNNGLIDAFALSSGN	480

Db 67 TKRVYERNVSIIPMTWTKSEYFIPKQESYDQADVIVANPYLKYGDPTLQYGRCE 126
 Qy 128 KGERIHTLPDFIACKKIAEYGPQGRVHWAHLRWCPDEYNNDEKEYLS-NGR10A0VR 186
 Db 127 KGKXIHFTNPNHIIYSGRGRVHWAHLRWIFDEYNNVQDFYISRKNTIATR 186
 Qy 187 CSAGITGTNVV-KKCGSGCYTKKCTFNKVTGLYKCECFVLOSROTEKASIMFAQHVD 245
 Db 187 CSTHTIIGINVFKKPGGSCITSLCRDSQIGLYEAKCTFLPKKQSAKESIMFMSLHS 246
 Qy 246 IVEFTEQHNNKEAPNKNQCNLRSTWVIRDSDEKKTTPMT--TQPNPPTSLLOIG 303
 Db 247 VTEFCTEHTNTEAPNLQNCNGKSTWDVIMNSVDFQNTSPMTMPPHTPFTSLKSK 306
 Qy 304 QRIVCLVLDKSGMATNRLNQAOLFLLOTVELGSGVMGTVDFAAHVQSELTOIN 363
 Db 307 QRVVCVLDKSGMGAEDRLFQNNQAELLYLQVIEKSLVGMVTFDVAEIQHLTRIT 366
 Qy 364 SGSDRDTLAKRLPAAASGGTSCSLRSF-TVIRKKYPTDGSSEIVLLTDCEDNTISGCF 422
 Db 367 DDNVYQKITAKLPQVANGGTSICRGLKAGFQAIHSDOSTSGSEIILLTDCEDNEINSCF 426
 Qy 423 NEVKQSAIHTVALGPSAAOELKMTGGLQTYASDVQNNGLIDAFALSGNGAV 482
 Db 427 EDVKRSGLIHTALGPSAAKELETLSNMTGVRPFANKDI--TGLTNAFSRISSSGSI 484
 Qy 483 SQRSIQLESKGLTQNSQWNGTIVVDSTVGDFTFLITMTTQPPQIILLWDPG--KQG 540
 Db 485 TQQAIOLESKALKITGKRVNGTVPVDSTVGNDFVFWTWTIQKPEIVLQDPKGYKTS 544
 Qy 541 GFVYDK-NTKWAYLQIPGIAKVGTKYSL---QASSQTLTLTVTSRASNATLPPITVTISK 596
 Db 545 DFEDKLNIRISARLQIECIAETGTWTYSLNNHASSOMLTVTVTTRSPPTIPPVIATAH 604
 Qy 597 TNKDTSPFPLVYVANIROGASPIRLASTALIESVNGKVTLELLDNGAGADATKDDG 656
 Db 605 MSQHTAHPSPMIVYVQVSGFLVGLISVIAIETEDGHQVTLLELDWNGAGRDYKNDG 664
 Qy 657 VYSRYETDYNRGYSVKVRALGVNAARRVIPPQSGALYIPGWIENDEIQNPPRPEI 716
 Db 665 IYSRFTDYGNGRYSKLVHAQARNNTARLNROPKNKLYVGYVENGKRIILNPPRPEV 724
 Qy 717 NKDDYQHKOVCFSTSGSGSVASDV-PNAPIDLPFGQITDLKAEIHGGSILNLTWTA 775
 Db 725 KDLAKAKIEDFSRLTSGGSFTVSGAPPGNHPSVFPFPPSKITDLEAKFK-EDYIQLSWTA 783
 Qy 776 PGDDYDHGTFAHKYIIRISTSLDLRDKENESLQVNTTALIPKEANSSEVLFKPEITFE 835
 Db 784 PGNVLDKGNKANSYIIRISKSFMDROEDFDNATLNTSLNIPKEAGSKENTFEKPEHFRVE 843
 Qy 836 NGTDLFIATQAVDKVLDKSEISNIARVSLFIP 867
 Db 844 NGTKFVIVQVAINELISEVSHIVQAIKFIIP 875

RESULT 3

S76691

hypothetical protein - *Synechocystis* sp. (strain PCC 6803)

C:Species: *Synechocystis* sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S76691

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocys*

s.

A:Reference number: S74322; MUID:97061201

A:Accession: S76691

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-420 <KAN>

Qy 480 GAVSORSTOLESKGLTLQNSOMNGTIVVDSTVGKDTFLITWTTPQIILLWDPGQKQ 539
 Db 481 AAIQHSIQLESRGVNLQNNMGNSVVDSSVGKDTFLITWTTHPTTIFWDPGVEQ 540
 Qy 540 GGFVVDKNTKMAIYLPQIAKVGTKYSLOASSOTLTLTVTSRASNATLPPITVTSKNK 599
 Db 541 NGFILDITTKVAYLPQGTAKVGFWKYSIQASSOTLTLTVTSRASATLPPITVTPVVK 600
 Qy 600 DTSKSPVLVVANTROGASPIRLASTALIESVNGKVTLELLDNGAGADATKDDGYVS 659
 Db 601 NTGKFPSTVTVASITROGASPIRLASTALIESVNGKVTLELLDNGAGADATKDDGYVS 660
 Qy 660 RYFTTDTNIGRYSVKVRALGVNAARRVIPPQSGALYIPGWIENDEIQNPPRPEINKD 719
 Db 661 RFTTADANGRYSVKTKWALGGVTSORAAAPPKNRAMIDGWIEDGEVYRMPNPPRPTSY- 719
 Qy 720 DVQHKOVCFSTSGSGSVASDVP-NAPIDLPFGQITDLKAEIHGGSILNLTWAPGD 778
 Db 720 -VQKQCFSTRSGSGSVATNVPAAPIDLPFGQITDLKASIQGNLNLNLTWAPGD 778
 Qy 779 DYHGTAKHYIIRISTSLDLRKFNSLOVNTTALIPKEANSSEVLFKPEITFENG 838
 Db 779 DYHGRASNTIIRMTSIVDLRHFNTSLQVNTTGLIPKEANSSEVLFKPEITFENG 838
 Qy 839 DLETAQVADKVDLKSISNIARVSLFIPQTPPTSPDTSAPCPNTHINSTIPGHI 898
 Db 839 DIFTAQVADKSNLKSISNIARVSLFIPQTPPTSPDTSAPCPNTHINSTIPGHI 898
 Qy 899 LKIMKWKIGELQLSI 913
 Db 896 LKIMKWKIGEMQVTL 910

RESULT 2

T02205

Lu-ECAM-1 protein - bovine

C:Species: *Bos primigenius* taurus (cattle)

C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 01-Dec-2000

C:Accession: T02152; T02151

R:Elble, R.C.; Widom, J.; Gruber, A.D.; Abdel-Ghany, M.; Levine, R.; Goodwin, A.; Pauli,

submitted to the EMBL Data Library, April 1997

A:Description: Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial ch

A:Reference number: Z14590

A:Accession: T02205

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-905 <ELB>

A:Cross-references: EMBL:AF001261; NID:g2623762; PIDN:AAB86529.1; PID:g2623763

A:Experimental source: lung

A:Accession: T02152

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-793, 'SGSPF', 'M', '862', 'RF', '865-866', 'Q', '868', 'AKVLEIQ', '876', 'QHQ', '880', 'FQ', '882

A:Cross-references: EMBL:AF001263; NID:g2623766; PIDN:AAB86531.1; PID:g2623767

A:Experimental source: lung

A:Accession: T02171

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-792, 'ES', 'EL3>

A:Cross-references: EMBL:AF001262; NID:g2623764; PIDN:AAB86530.1; PID:g2623765

A:Experimental source: lung

Query Match 49.08; Score 2334; DB 2; Length 905;

Best Local Similarity 53.4%; Pred. No. 6e-146;

Matches 466; Conservative 144; Mismatches 246; Indels 16; Gaps 11;

Qy 8 VFILHLLEGALNSLIQLNNNGYEGIVVAIDPNVPEDTLTQIKDMVQTASLYLFEA 67

Db 8 ILFLHLPLPG-MKSSMWNLINNGYDGVIAINSPVDEKLEIKEMVTEASTLYLFA 66

Qy 68 TGKRFYFNKVAIIPETWTKADYVRPKLETYKNADVIAESTPPGNDEPYTEQMNGCE 127

Qy 230 RTEKASIMFAQHVDISIVFECTEQNHUNKEAPNKQNKCNLNSTWEVIDRSEDFKKTP-- 287
| | | : | : || : || : || : ||
Db 177 SOLSGGGIEGG---STIF-----SGINQATFSSNTAEVVPF----ETTPNP 216

Query Match
3.2%; Score 153; DB 2; Length 689;

Best Local Similarity 26.6%; Pred. No. 0.042; Matches 68; Conservative 43; Mismatches 97; Indels 48; Gaps 12;	
QY 310 VLDKSGMATGNRLNQAQOLFLLQTVELGSGWGMVTFDSAHHVQSELQINSQSDRDLAKRLPAAASGCTSIKSGL 389	QY 330 GOLFLLOTVELGSGWGMVTFDSAHHVQSELQINSQSDRDLAKRLPAAASGCTSIKSGL 389
Db 254 VLDISGMA-GTKALLKRA-MGFVIONLGSNDRLSVLAFSTARRLLPLTKMSDAGRQ 311	Db 520 QSMVVQAFTLSQ-----KSSVSLSSQTLSDSHS---TATLTFIAHDAAGNPVI-GL 569
QY 370 TLAKRLPAAASGCTSIKSGLSAFTVI---RKKYPTDGESEIVLLTDGEDN-----SAA 442	QY 390 RSAFTVIRKKYPTDGESEIVLLTDGEDNITSGCFNEVKQSGALIHHTVALGP-----SAA 442
Db 312 ALQAVNSVVANGTNAEGLRGKGVKVMEDRRDKNPV--ASIIILSDGRDITTMQADPNY 369	Db 570 -----VLSTRH--EGVODITLSDWKDNG--DGSYTOILLTGAMSGTLTLMPLNGVDAAKA 621
QY 417 -----TISGCFNEVKQSGALIHHTVALGPS-AAQELLESKMTGCLQTV-ASDQVONNGL 468	QY 443 QELEELSKMTGGTQTVASQVQVONNGLIDAFGALSNGAVSQRSIOLESKGLTLONSQNM 502
Db 370 KLLPLSMHGC--ESKRFOIPVHSFGFGDHDASLMHVSFSGTFSFIESVIODAL 427	Db 622 PAVNIISVSSS-RTHSSIKIDKDRYL-----SGNPIEVTVELRDENDKPKVQKQOL 673
QY 469 IDAFGALSNGAVSQRSIOLESKGL-----TLQNSQWNGTV-IVDSTV 512	QY 503 NGTVIVD-----STVGKDT--LFLITWTQ-----PQILL--WDPGQKQGGFVVD 545
Db 428 AQCIIGLLS-----VAVQELRLIEGMCSDVHLSSIKAGSYQSLVSGDGRSGCVDIGDLYA 483	Db 674 NTAVSIDNVRPGVTDDMKETADGVYKATYATYTKGSLTAKLLMQNWN-EDLHTAGFIID 732
QY 513 GKDTLFLITWTQPPQ 528	QY 546 KNTKMAVLOIPGIAKVTGKYS-----LQASSOTILTITVTSRASNATLPII----- 591
Db 484 DEERDELIS-VNIPPO 498	Db 733 AN-----PQSAKIATLSASNGVLANENAANTVSVNVADEGSN---PINDHTVTFA 780
RESULT 6	QY 592 ---TVTSKTNKDTSK-----FPSPLVV----- 610
E85822	Db 781 VLGSATSFNNQNTAKTDVNGLATFDLKSSKQEDNTVEVTLENGVKQTLIVSVFGSSSTA 840
probable invasin Z3135 [imported] - Escherichia coli (strain 0157:H7)	QY 611 YANIROGASPIL-----RASVTALIESVNGK-----TVTLELLDNGAGA-----DATKDD 655
C:Species: Escherichia coli	Db 841 QVDLOKSKNEVADGNSATWATVRODAKGNLLNDVAVTENV-NSAAKLSQTEVNSHD 898
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001	QY 656 GVSRYFTYDTNGRYSYKVALGGVNAARRVI---PQSGALYIPGWIENDEIQWNP 712
C:Accession: E85922	Db 839 GIATATLTL-KNGDVTYTVASVSG-SQANQOVIFIGDQSTAALTLS--VPSGDITVTNT 954
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew	QY 713 RP-----EINKDDVQHKQVCS-----RTSSGGSFVAS----- 740
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,	Db 955 AFLHMTATLQDKNGNPLKDKKEITFVSPNDVASRFSINSKGKMTDSNGTASLTGTLAG 1014
Nature 409, 529-533, 2001	QY 741 -----DVPNAPIDLPFPQGITD-----LKAHIGGSLINTWTAPGDDYDHGT 784
A:Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.	Db 1015 THMITARLANSVSDTQPMTFVADKRAVVVLOVTSKAEIINGVDETTLTATVTKDPFDNV 1074
A:Reference number: A85480; MUID:21074935; PMID:11206551	QY 785 AIKYIIRISTSLDLRDKFNESLQVNTTALIPKEANSEVEFLFKPENITFENGDLFIAT 844
A:Accession: E85922	Db 1075 VRNLSVVFRTSPADTQSLN-ARNTNENGIAEAVTLKGVLGVHTAEALLNGNRDTKI-- 1131
A:Molecule type: DNA	QY 845 QAVDKVDLKSEISNIARVSLFIPQTPETPS-----PDSETSAPCNHINSTIP 894
A:Residues: 1-2660 <STO>	Db 1132 -----VNIAPDASN-AQVTLNIPAAQVVTNNSDSVQLTATVKDPSNHPVAGITVNTMP 1184
A:Cross-references: GB:AE005174; NID:g12516151; PIDN:AAG57041.1; GSPDB:GN00145; UNGP:231	RESULT 7
A:Experimental source: strain 0157:H7, substrain EDL933	C81265
C:Genetics:	probable lipoprotein Cj1678 [imported] - Campylobacter jejuni (strain NCTC 11168)
A:Gene: Z3135	C:Species: Campylobacter jejuni
Query Match 3.2%; Score 152; DB 2: Length 2660;	C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
Best Local Similarity 18.4%; Pred. No. 0.43;	C:Accession: C81265
Matches 199; Conservative 163; Mismatches 385; Indels 332; Gaps 55;	R:Parkhill, J.; Wren, B.W.; Muddall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chli
QY 71 RYFKNVALLIPETWTK---ADYVRPKLEYKNADVLVAESTPPGNDPEYTEQMNGCE 127	C.W.; Quail, M.; Rajadream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba
Db 183 REGTARITLGVDEFSKNSOFDLHPWYETPDN--LFFSQHTLHRTDE--RTQINN--G 236	Nature 403, 665-668, 2000
QY 128 KGERIHLTPDFIAGKLAFCYCGRAFVHEWA--HLRWGVDFEYNNDEKFLYLSNG--RIQ 183	A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
Db 237 LGWR-HFTFTWNGINF-----FFDHLRSYHSRAGIAGYWRDYDKLSSNGYLRLT 287	A:Reference number: A81250; MUID:20150912
QY 184 AVRCSAGITGNVKKQCG-----GSCYTKRCTFNKVTLGYEKGEFVLQ 228	A:Accession: C81265
Db 288 NWRSAPELNDYEAPANGWDYRAEGWLPANPHLGKGLVIEQYIGDEVALFDK-----D 341	A:Status: preliminary
QY 229 SRQTEKASTMFAQHYD--SIVFECTEQNHKCAPNKQKCNLRLSTWE----- 274	A:Molecule type: DNA
Db 342 DROSNPHTATAGLNTYTPFLMTFSAEORQKQGEN--DTRFAVDFTWQPGSAMQKOLDPN 399	A:Residues: 1-928 <PAR>
QY 275 -----VIRDE---DPKK-----TTPMT--TOPPNPTFSLLOI----- 302	A:Cross-references: GB:AL139079; GB:AL111168; NID:g96968971; PIDN:CAB73665.1; PID:g9696
Db 400 EVDARRSLAGSRDLVDRNNNIVLYRKKELVRLTLTDPVTGKSGEVKSLVSLQTKVAL 459	A:Experimental source: serotype O2, strain NCTC 11168
QY 303 -GORIVCLVLKSGS--MATGNRL-----NRLNQA 329	C:Genetics:
Db 460 KGYNVEATALEAAGGKVVTGDKDILTLPAYRFTSTPDTNTWPIETVAEDVKGNFSNRE 519	A:Gene: Cj1678
	Query Match 3.1%; Score 145.5; DB 2: Length 928;
	Best Local Similarity 18.6%; Pred. No. 0.21;
	Matches 157; Conservative 100; Mismatches 261; Indels 327; Gaps 39;

Qy	150	QGRFVHEWAHLRWG--VFDEYNDEKFXLSNGRIQ-----AVRCSAGITGTNVYKCKQG	203
Db	110	ENTGIIH--AEGEWGSSILLEYGCFIHIINTGITSNNNVGIGSAYGVFGTIIK--DGG	165
Qy	204	SCYTK-----RCTFNKVTGLYKGCCEFYVLOSROTEKASIMFAQHYD	244
Db	166	MVYKYSNAIGVRSQTLGDLYIDGRSNGTVSGIYSEEHGILLEN-----	210
Qy	245	SIVEFCTEQHNKNEAPKNQKCNLRKSTWEIVRDSDEKTKTTPMTTPPPNPTFSLIQIGQ	304
Db	211	-----NSRTOKIELKN--GGIIKGNID-----GI	232
Qy	305	RIVC-----LYLDKSGSMATGNR--LNRLNOAGOLFLLQTVELGSMV-----	344
Db	233	KLINSASLUSGEMILSGEGRVGEGRGVGIIILNRSGKIEGSIKVEDGATVTTATSNRAIANSG	292
Qy	345	-GMVT-----FDSAHHVQSEL-----IQIN	363
Db	293	SGSITGGITVSGKNTKLEGNIINTGNASIGSDIKIEGGAKEVGGLYNOGNGSISGSVQVS	352
Qy	364	SGSRDRLAKRLPAASGGTSGIC--SGLRSAFTVIRKKYPTDGSSEIVLLTGDENPTISGC	421
Db	353	GGSSIDITNEGNAISGSIIVYKDKSLDITSTSSIGSGS-----ITNNSDNKL-----	405
Qy	422	FNEYKQSGAIHITVALGPSAAQAEELSCKMTGGL--QTVASDOV---QNGGLIDAFGALS	476
Db	406	-EISNSGNI-----GKIESTGSADMWISNMGGTISG--GISS	441
Qy	477	SGNAVQSRSIQLESKGLTLQNSOMMNGTVIVD---STVGKDTLFTLTTTQPPQILW	532
Db	442	SGSGSTSISS---SGGSTINNGITVSGSAQVEISNOGSGVKD-----	480
Qy	533	DPSCKQKGGFVVDKNTKMAVLIQPTAKVG--TWKYSLQASSQTLTLVT--SRASNATL	588
Db	481	-----ENGNTVTNN-----GSGSVGIKDWLSTDKNTKGLNVTVVGSRAFNFKV	525
Qy	589	PPITVTSKTNKDTSKFPSPLVVYANIRQASPIRLASVTALIESVNGKVTVLELIDNGAG	648
Db	526	ENITV-DQSNVDLEEL-----NDNINISGVNQNNIG-NIGTINGSG	564
Qy	649	ADATKDGVSRYSTFTTVDYTNGRYS--VKVRALGVGNAAARRVYIPQSGALYIPGWIENTEI	707
Db	565	EISLFPDIPITGKLTDFNLNASISGATFRSLISITTSRRSTFI-----DHW	609
Qy	708	QWNPPEINKDDVQHQKQVCFSTSSGGSFVADVPN--APIDLPFPQGITDLKAEIHG	765
Db	610	MGN-----SMQSFALASSKKSQSITAMSEKGNLYADASY-----IKSDLAN	650
Qy	766	GSLNLTWTAPGDDYDHGHTAHKYIIRISTGILDRKFNESLQVNTALI-----PKEAN	820
Db	651	GS-----YGSNKHESFLIPLPYSQNVLSLNEBSKGHTKGTIIIGYSTLKLKDSG	698
Qy	821	SEEVF-----LFKPENITF-----ENGTDLFTAIO--AVDKVOLK	853
Db	699	IYGVAGYEDTKMGSTVFDINNRITYAGLYKTYNLTFTTEKQGEVYIINAQKKAALIKNDLT	758
Qy	854	SEISN	858
Db	759	EKIGN	763

RESULTS

RESULT 8
 B85547
 probable rpx family exoprotein [Imported] - Escherichia coli (strain O157:H7)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: B85547
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of
 A:Reference number: B85480; MUID:21074935; PMID:11206551

E;Accession: B85547
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-5188
A;Cross-references:
A;Experimental source:
C;Genetics:
A;Gene: Z0615

Query: Match					
Best Local Similarity 3.0%; Score 144.5; DB 2; Length 5188;					
Matches 117; Conservative 69; Mismatches 204; Indels 153; Gaps					
Qy	314	SGSMATGNRLNRINQAGQLFLLOTVELGSWV-----CMVTFF-----DAAH- 354	: : : : : : : : : : : : : : : : : : : :		
Dd	3362	TGAVA-GDLR-VVTIAGQQYVTTSDASGNWSGVGPASVISGLADGTVIISATIYDSAGNS 3419	: : : : : : : : : : : : : : : : : : : :		
Qy	355	-VQSELTIQINGSS-----DRDTIAKLRLPAAASGGTSCIGLSAFSTVIRKKYPYTDGS 405	: : : : : : : : : : : : : : : : : : : :		
Dd	3420	STOPIHNQVVNTAAVSLSVSTISSDNL--INAAGAQSALTLSGTCTNFAT-----GT 3468	: : : : : : : : : : : : : : : : : : : :		
Qy	406	EIVLLTDGEDNTISGCIFNEVKVGSAIIHTVALGPSAAQLEELS KWTGLQTYASDOVN 465	: : : : : : : : : : : : : : : : : : : :		
Dd	3469	VTVLLNGK-----GISATIQSGSVMNVXADVA-----LSDGTSYTVSASAQD 3515	: : : : : : : : : : : : : : : : : : : :		
Qy	466	NGLIDAFALSSGNGAVSORSIQLESKGLTLNOSMMNGTIVJDSVTGCKDTLFITWTTQ 525	: : : : : : : : : : : : : : : : : : : :		
Dd	3516	-----SAGNSATASRAVD-----LTAPVISINTVTDTRXNAAEQQQ 3554	: : : : : : : : : : : : : : : : : : : :		
Qy	526	PQILLMDPSQGKGFEVDKNTKMAYLOIPGIKVGIWKYSLOA-----SSQILTL 577	: : : : : : : : : : : : : : : : : : : :		
Dd	3555	P--LTLNGSTSAEYGOTVTVTFGCKTYTAT--VAANGTWALKNPVAVDLAALGCGAQTTTA 3610	: : : : : : : : : : : : : : : : : : : :		
Qy	578	TVTSRASN-----ATLPPIVTSKTNKDTSKFSPVLVYVANIROGASPILRASV 626	: : : : : : : : : : : : : : : : : : : :		
Dd	3611	SVNDRACNPGOATHALTVDTPVAPTIVTAVAGDD-----IINNAEQLAGOTISGTT 3661	: : : : : : : : : : : : : : : : : : : :		
Qy	627	TALIESVNGKYTVILLELDNGAGADATKDGGVSRVFTHDYDNCRYSVKVRA--LCGVNAA 684	: : : : : : : : : : : : : : : : : : : :		
Dd	3662	TAEV---GQTIVTF--NGQWTSATVSGS-----GSWSVIPAQCFAGUSDG 3703	: : : : : : : : : : : : : : : : : : : :		
Qy	635	RRRIPOOGSALXPIMWIND-EIOWNNPPRPINK---DDVOHKVCFSRTSSSGSFVAS 740	: : : : : : : : : : : : : : : : : : : :		
Dd	3704	SYTISATVSDOAGNPCSASRGVTLNGDVPTVTINTFAGDDV-----VNAAEHGSSLVIS 3757	: : : : : : : : : : : : : : : : : : : :		
Qy	741	DVPNAPIPDLEFPQCITDLKAETHGSLNLNLTWPAGDDYDHGTA-----HKYIIRI 792	: : : : : : : : : : : : : : : : : : : :		
Dd	3758	GTTTAPV-----GQ--TLPLTLNGKTYTTTTVOTGGWSYTLGSDADV TALADGNAYVINA 3809	: : : : : : : : : : : : : : : : : : : :		
Cy	793	STS 795			
Dd	3810	SVS 3812			

RESULT . 9

RESULT : 9
 371704.
 cell surface antigen (sca3) RP451 - Rickettsia prowazekii
 C:Species: Rickettsia prowazekii
 C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
 C:Accession: B71704
 R:Andersson, S.G.B.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark
 Nature 396, 133-140, 1998
 A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
 A:Reference number: A71630; MUID:99039499
 A:Accession: B71704
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2340 <AND>
 A:Cross-references: GB:AJ235271; GB:AJ235269; NID:G3868717; PID:G386
 A:Experimental source: strain Madrid E
 C:Genetics:
 A:Gene: sca3; RP451

Query Match 3.0%; Score 143.5; DB 2; Length 2340;
Best Local Similarity 20.0%; Pred. No. 1.3;
Matches 162; Conservative 95; Mismatches 291; Indels 261; Gaps 38;

QY 164 GFVEYNNDEKFLYSLNGRI--QAVRCSAGITGTWVKKCGGSGYTRKCFNKVTGLYEK 221
DB 840 GDIDFNKAGKFLGDGAMIDGVLNCGVAGT----- 872
QY 222 GCEFLVLOSROTEKASIMFAQHVDSIVEFCTEQNHKNKAPNOKNOKNLRSTWEYRSED 281
DB 873 -LDFIGDGNVNTQIGADNANSISTI-----NIQDNTK-----NVTIANDIFVDNIH 918
QY 282 FKTTPTMTTPPPNPTFLSIQIQRIVCLVLKSGSMAT-----GNRLNRLNQA---GQLFL 334
DB 919 FTNG-----GILQGLGNTLTHNIDFGANGGTLEFNGNNTYLNLAIIVNGQGI 966
QY 335 LQTV-----ELSGWGWMTFDSNAHVQSELIQINSRDRDTLAKRLPAAASGGTSIC 386
DB 967 LNAFTNLKASDDTIGT-VKIIINGIQTPTQNTIOVNN-----KNLTIVSSVNSIN 1017
QY 387 SG-----LRSAPVIRKPKPTDGTSEIVLLTDGEDNIVISGCFNEVKOSGALIHV 435
DB 1018 FGDANSQILSAPVDQTIKFNNLNETGGIITLDSGNLNTISG-----NNG-----I 1066
QY 436 ALGPSAAQAELELSKMTGGLOTYASD-QVONNGLIDAFGALSSGNGAVSORSIQLESKGL 494
DB 1067 KLG-SKGNELSSLN--IRGKVVTVDLDION-----IHLNINNGAL-----FDDQSL 1111
QY 495 TLQNSOMN--GTIVDSVGVKOTLFLITWTTPPOILLWDPGKOGGVVDKNTK--MAY 552
DB 1112 TSAKININIGTVAGGATYLDAL-----NDNF--DLNTSGMVF 1148
QY 553 LQIPGIAKVTKYSLQASSQTLTLVTSRASNATLPPITVTSKTNK----- 599
DB 1149 KHDSILEL---KNSSNTDHTITLSALDPGNQGTGIIKLTIDTKLIDNNGNVAYTL 1205
QY 600 DTSKFPPLVYANIRGASPIILRASVTALIESVNGKTVTLELD----- 644
DB 1206 GTANHMLKOLTFASIDNGA-----IALKVGINVENVTLNKIDELNEVANVLFNKN 1257
QY 645 -----NG-----AGADATKDD-GVYSRYFTYDNGRYSVKVRALGVNAARR 687
DB 1258 TTYTATGNINGHVDFOGNAGVINLNDIEIDGVSITSTGNVNG--TLNFGSGKVTGLINN 1315
QY 688 VLPQSGA---LVIPGWIENDEIQWPPR-----PEINKDDVQHKQVCFSRYS 732
DB 1316 IVMLQAGAGDVLSASGNSYITELOGNNNLTFAANSHLTDTINKTGGQDLNLF---I 1372
QY 733 SGGSFVADVPNAPIDLPFGQITDLKAEIHGSLI-----NLWT-APGDDY 780
DB 1373 NGGSVSGSIGANAAGVDIIINAGSVNFSNTLKSQNVISDGATMQVNNVTATDISKNA 1432
QY 781 DHTAHKYIIRISTILDRKFNESLQVNTALIPKE-----ANSEEVF--LFKPE 830
DB 1433 NNGTL-----KLNNHTPINTSLGNNAIGTIEVANNDVTIIGTLQAO 1476
QY 831 NITFENGDLFLTAIOAVDKVLDKSEISNI 859
DB 1477 NIHFSNATQ-----AATLTGAAASQVTNI 1500

RESULT 10
T21810
hypothetical protein F35G12.4b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21810
R:Chui, C.
submitted to the EMBL Data Library, October 1994
A:Reference number: Z19473
A:Molecule type: DNA
A:Accession: T21810

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA
A:Residues: 1-683 <WIL>
A:Cross-references: EMBL:Z46242; PIDN:CAA86338.1; GSPDB:GN00021; CESP:F35G12.4b
A:Experimental source: Clone F35G12
C:Genetics:
A:Gene: CESP:F35G12.4b
A:Map position: 3
A:Introns: 62/3; 105/1; 135/3; 200/3; 269/3; 332/2; 390/1; 425/3; 458/3; 490/1; 542/2

Query Match 2.9%; Score 138; DB 2; Length 683;
Best Local Similarity 20.6%; Pred. No. 0.4;
Matches 124; Conservative 85; Mismatches 191; Indels 202; Gaps 33;
QY 416 NTISGCFNEVKOSGALIHVVALGPSAAQEL-----EELSKMTGGLOTYA 459
DB 172 NNIGC---KDS---IYSLATTPNLVLGAGTEKICRLFDPRTEKIMKRLRHTDNVR 224
QY 460 SDQVQNNGLIDAFGALSSNGAV-----SQRSIOLESKGLTLQNSOMNCTVIYDS-- 510
DB 225 ALVNDGDT---RAISAGSDATIRLWDICQORCI---ATCIAHEGVW---TLQVDSSE 274
QY 511 ---TVGKDFLEIT--WTQPPQIILLMDPSGQKQGGVVDKNTKMAYLQIPGIAKVGM 564
DB 275 TTVYSAGKDKMVKTPLYDFTKQLLKEAPVKK-----LLESEKDNPVSLWVGTM 326
QY 565 K-----YSLOASSQTLTLVTSRASNATLPPITVTSKNTKDTSKFPPLVYVANIROGAS 619
DB 327 KSDIKRWSIRPSAQ---LSIGGDEGDS-----TSNANHSVSASSPPVTFKYIR--- 373
QY 620 PILRASVTALIESVNGKTVTLELDNGAGADATKDDGVYS--RYFTYDNGRYSVKVR 677
DB 374 -----VKDQKGOOSTPELVI--PGAPAIKKHAKLSKRVLTRDSGNNV---A 416
QY 678 LGVNAARR-----RVIPQSGALYIPGW-----IENDEI---QWNP 712
DB 417 LYDLAARKIKDYGRIFEYVDNSRQYIPSVFVVDKSGMLQITLDELDSLWSLSS 476
QY 713 RPE-INKDDVQHKQVCFSRYSSTSGSFVADVPNAPIDLPFGQITDLKAEIHGSLINL 771
DB 477 KDAGFDDNDRETQII---VNYGGMLRS-----LFEWPPCKMTNVDA----- 517
QY 772 TWTPAGDDYDHGTAHYIIRISTILDRKFNESLQVNTALIPKEANSEEVF--LFKP 829
DB 518 ---ADADDVQKAT-----LNFISLPEH-----TPLIICEGNGRPLRLVGD 556
QY 830 ENITEE-NGTDLFLTAIOAVDKVLDKSEISNIARVSLFI--PPQTPPETPSPDSTA--- 882
DB 557 AGKEFEANELAQIAPMWVIDAIE-RNQLPKFNKMPFYLLPHPSINPKPKKDRLSATEML 615
QY 883 -----PCPNIIHINSTI-----PGIH---ILKIMKWKIGEL 909
DB 616 QVYKVMHVEYKILSTNDVGSIPLNQIHTKWMYCNQORLEPMDLRTVKHLYNKQSGEL 675
QY 910 QL 911
DB 676 LL 677

RESULT 11
T21808
hypothetical protein F35G12.4a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21808
R:Chui, C.
submitted to the EMBL Data Library, October 1994
A:Reference number: Z19473
A:Accession: T21808
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-686 <WIL>

A:Cross-references: EMBL:Z46242; PIDN:CAA86335.1; GSPDB:GN00021; CESP:F35G12.4a

A:Experimental source: clone F35G12

C:Genetics:

A:Gene: CESP:F35G12.4a

A:Map position: 3

A:Introns: 62/3; 105/1; 135/3; 200/3; 269/3; 332/2; 390/1; 425/3; 458/3; 490/1; 542/2; 6

Query Match 2.9%; Score 136.5; DB 2; Length 686;

Best Local Similarity 21.1%; Pred. No. 0.51;

Matches 114; Conservative 81; Mismatches 176; Indels 165; Gaps 30;

QY 416 NTISGCFNEVKSGAIIHTVALGPSAAQEL-----BELSKMTGGIQTYA 459

DB 172 NNLIGC-----KDS---IYSIATTPNLSVLGAGTEKIRLFDPTNKIKMLRHTDNVR 224

QY 460 SDQVQNGLIDAFGALSSNGAV-----SORSIQLESKGLTLQNSQWNGTVIVDS-- 510

DB 225 ALVNDGDT---RALSAGSDATIRLWDIGQORCI---ATCIAHEEGVW---TLQVDSSE 274

QY 511 ----TVGKDTLFLIT--WTTQPPQILLWDPGSGQGGFVVDKNTKMYLQIPGIKAVGTW 564

DB 275 TTYSAGKDKMVKTPLYDFTKSOLLKEEAPVKK-----LLLSEKDNVPVSLWYGTW 326

QY 565 K-----YSLQASQTLTLVTSTRASNATLPPITVTSKTNKDTSKFPSPLYVYANIROGAS 619

DB 327 KSDIKRSIRPSAO---LSGGDEGDS-----TSNANHSVASSSPPVTFKYIR----- 373

QY 620 PILRASVTALIESYNGKVTLELLDNGAGADATKDDGVS--RYFTYDTNGRYSVKVRA 677

DB 374 -----VKDQKGQOSTPELVI--PGAPAIIKHAMLSDKRHVLRDSDGNV-----A 416

QY 678 LGGVNAARR-----RVIPOOSGALYIPGW-----IENDEI-----QWNP 712

DB 417 LYDVLAARKINDXGKRIFEEVVDENSQVYIPSWFVVDKSGMLQITLDELDAISSWLS 476

QY 713 RPE--INKDDVQHKVCFSRSSGGSFVSDVPNAPIDLPFGQITDLKAEIHGSLINL 771

DB 477 KDAGFDNDRETKOII---VNYGGMURS-----LFRPPCKMTNVDA----- 517

QY 772 TWTAPGDDYDHGTAHKYIIRISTSLDRKFNESLOVNTTALIPKEANSEEVF--LFKP 829

DB 518 ---ADADDVQKAT-----LNFISLPEH-----TPLICEGNGRPLYRLVGD 556

QY 830 ENITFE--NGTDLFTAIQAVDKVLKSEISNIARVSLFI--PPQTPPETPSPDETSA 882

DB 557 AGREFEANELAQIAPMWVIDAIE--RNQLPKFNKMPFYLLPHPSTNPKQPKDRLSA 611

RESULT 12

A85547

hypothetical protein Z0609 [Imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001

C:Accession: A85547

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: A85547

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1461 <STO>

A:Cross-references: GB:A80051174; NID:g12513364; PIDN:AAG54837.1; GSPDB:GN00145; UWGP:Z06

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: Z0609

Query Match

Best Local Similarity 21.1%; Pred. No. 1.7;

Matches 160; Conservative 95; Mismatches 282; Indels 221; Gaps 40;

QY 290 TOPNPFTS-----LLQI-----GQRIVCLVLKSG-----SMATGNRLNRLNQAQQLFL 334

DB 157 TDDTPTFGSGQAGATIQIKDSNGNTIASTQVDNNGHWSVSLPT-----QSAGEHTW 249

QY 335 LQTVELGSWV---GMVTF---DSAAHVQSELIQINSGSDRDTLAKRUPAAASGGTS--IC 386

DB 250 SVVOIVGSTTIDAGSIITLTDNSQASVQ---VATTAGNITINASEQAAGFTLSGTSLSHA 306

QY 387 SGLRSATFVIRKKYPT-----DGSEIVLLT--DGEDNTISG--- 420

DB 307 QGTETLTNLGKTVTTSVVGANGAMSVQVPTADAQALBEGNQAVLVSGKDATGNTVTGAQL 366

QY 421 -----CFNEVKSGAIL---HTVAL---GPSAAQAELEELSKMTGG---LQTVASD 461

DB 367 LTVDTPPTLAINITIAQDNIIISAAEHNVALVLSGTSNAEAGQTVTLTVNGKSHSTATVGS 426

QY 462 -----OVQNGLIDAFGLSSGNGAVSORSIQLESKGLTLQNSQWNGT-----VI 507

DB 427 GTWQVTLTPATEVQ-----ALAEQYAVN--ASVSDRAGNTTSHSANFVDTISAPVVS 476

QY 508 VDSVVGKDTLFLITWTPPQILLWDPGSGQGGFVVDKNT--KMYLQIPGIKAV--GTWK 565

DB 477 VNTVAGDDIL-----NNAEQAVAQIIISGVSGASPGDVTVTKLGTHTVLTGIVLADGSMN 530

QY 566 YSLQAS-----SQTLTLVTSTRASNA-----TL-----PPIVTSKTNKDTSKPPS 606

DB 531 VALDPAVTRTLDRGANTIFVTVTDAGNTGAASRAITLVGVSPPLITINTVSGDD----- 584

QY 607 PLVYVANIROCASPILRASVTALIESYNGKVTLELLDNGAGADATKDDGVYSRYFTYD 666

DB 585 ---LIISAEKCAPLTLGSTQ---OXETGQIVTVTL-----AQGSFT-----TTVQ 624

QY 667 TNGRYSVKV--RALGGV-----NAARRVPIQQSGALYI--PGWIEN 704

DB 625 ADGWSLTVPAAMGNLPGDAVAITASVTLDSGNTGNTSRTITVDSQAPALSIDPLTADN 684

QY 705 --DEIQNPPPELNDKDDVQHKVCFSRSSGGSFVSDVPNAPIDLPFGQITDLKAE 762

DB 695 IINAAESGQDLPIGTGTTDAOPGQT--VTTLNGQTYQGVQPDGTSVTVPAANV----- 737

QY 763 IHGSLI--NLWTAPGDDYDHGTAHKYIIRISTSLDRKFNESLOVNTTAL-----I 815

DB 738 ---CALADGNATVTVASVNDVAGNPS-----SVSRVALVDATPPVVTINPVATDNVINT 787

QY 816 PKEANSEEVFLFKPENITFENGTLDTAIQAVDKVLKSEISNIARVSLFI----- 867

DB 788 PEHAQAQII---SGTVTGAQAGD--IVTVTLNNVDYTVTVVDGSGNWSLGVSPASVVSGLA 841

QY 868 ----PQTPPETPSPDETSA PCPNHIHINSTIPGIHLKI 901

DB 842 DGSYPVSVYVTDKAGNTGSQLSTVTVNTAAPLIGINSI 879

RESULT 13

A69222

hypothetical protein MTH911 - Methanobacterium thermoautotrophicum (strain Delta H)

C:Species: Methanobacterium thermoautotrophicum

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

C:Accession: A69222

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.

ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu

A:Reference number: A69000; MUID:98037514

A:Accession: A69222

A:Status: preliminary; nucleic acid sequence, not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-698 <MTH>

A:Cross-references: GB:AE000866; GB:AE000666; NID:g2621998; PIDN:AAB85409.1; PID:g262

A:Experimental source: strain Delta H

C:Genetics:

Query Match 2.9%; Score 136; DB 2; Length 1849;
Best Local Similarity 21.3%; Pred. No. 2.7;
Matches 204; Conservative 102; Mismatches 298; Indels 352; Gaps 53;

QY 109 STPPG--NDEPYEQ--MGNCG-----EKGERIHL--TPDFIAGKKLAIEYGPGRAPV 155
DB 279 SDPKGILSDPLTNYAVLGDGSGPLFVYDREKGMWLFISYDFWAGYNNKSW--QWNIYK 337

QY 156 HEWAHLRWGVDEYNNDEKFIYLSNGRIQAVRCASG--ITGTNVVKKCGQSCYKTKCT--F 212
DB 338 HEFA-----EKIY-----QOYSAGSLTGSN-----TOYTWOATGST 368

QY 213 NKVTGLYKGCCEFLVLOSROTEKASIMFAOHVDSIVEFCTEQHNHKEAPNKQKCNLR-- 270
DB 369 STITG-----GGEPLSDLTGKD-----KPNHGKSITLUGSGTTLNNH 408

QY 271 -----STWEVIRDSDFKTTPTMTTQPPNPTFS--LQIQGRIV 307
DB 409 IDGAGGLFEGDYEVKGTSDSTTWKAGVSVADGKT--VTWKVHNPKYDRLAKIGK-- 464

QY 308 CLVLD---KSGMATGNRLNRLNQAGQLLQ-----TVELGSWGVMTFDSAAHVQS 357
DB 465 TLVVEGKGKNEGLLVKGD-----GTVLKQKADANNKVQAFSQVGIVGSRST----- 511

QY 358 ELIQINSQSDRTLAKRLPAAAGGTSGCLRSATFVIRKKYPTDGESEIV--LLTDGED 415
DB 512 --LVLN--DKQVDPNSIYFGFRGLDLNGLSLTFDHRN--IDGARVVNNHMTNTSN 565

QY 416 NTISG-----CFNEVKQSGAIHHTVALGPSA 441
DB 566 ITITGESLITNPNTITSYNIEAODDDHPLRISIPYRPLYENODNRS---YYTLKKGAST 622

QY 442 AQELEELSKMT-----GGLQTVASQVON-----NGLIDAFGALSSGNGAVSORSI 487
DB 623 RSELPQSGESGENKMYMGRTSDEAKRNVNMHNNERNMNGENGVEFG--EETRATQNGKL 680

QY 488 OLESKGLTLQNSQWNGTVIV--DSTVGKDTLFLITWTTPQQLLLMDPSQKOGGFV-- 543
DB 681 NVTFNGSDQNRFLTGTGTNLNGDLNVEKGLFSLGRPTPHARDIAGISSTKDKPHFTEN 740

QY 514 -----VDKNTKMAVLOIPGIAKVTGWKYSLOASSOTLTLVTSRASN-----A 586
DB 741 NEVVVEDDWINRFRATMTNVGTNA-----SLYSGRNVANITSNITASNAQVHIGYK 793

QY 587 TLPPITVTSK-----TNKDTSKF--PSPLVVYANIROGASPL--RASVTALIES 632
DB 794 TGDTCVRSDDYGTGYVTCHNSLSEKALNSFNPTNLRGVNLTENASFTLGRKANLFGTIQS 853

QY 633 VNGKVTLE-----LLDNG-----AGADATKDDGVYSRYFTYDTNGRY 671
DB 854 IGTSQVNLAKENSHHMLTGSNVNQNLTNGHILHNAQDANK-----VTYNT----- 901

QY 672 SVKVRALGGVNAARRRVIPOQSGALYIPQWIEDEIQNPPPEINKDDVQHKQVCFST 731
DB 902 -LTVNSLSG-----NGSEFY--WVDFTNKSN--KVYVVK----- 931

QY 732 SSGGSFVSDVNPAPIDPLFPQGITDLKAEIHGSLI-----NLWTAPDDDYD 781
DB 932 SATGNFTL-----QVADKTGEPNHNELILFDASNATRNLEVTLANGSVD 976

QY 782 HGTFAHYIIRISITSLDLR-----DKFNESLOVNTTALIPKEANSEVFLKPEPNTFENG 837
DB 977 RG-AWKYKLRNVNGVYDLYNPEVEKRNQT--VDTT-----NITTND 1015

QY 838 TDLFIAIQAVDKVLDLKESEISNIARSLFPPQTP-----PETPSPDETSAP 883
DB 1016 -----IQA-DAPSQSNNEIARVETPPPPAPATESIAIEQPET--RPAETAQP 1063

RESULT 15
JC5575
inter-alpha-trypsin inhibitor heavy chain 2 - golden hamster
C:Species: Mesocricetus auratus (golden hamster)

A:Gene: MTH911

Query Match 2.9%; Score 136; DB 2; Length 698;
Best Local Similarity 20.3%; Pred. No. 0.57; 278; Indels 148; Gaps 28;
Matches 127; Conservative 74; Mismatches 28;

QY 161 LRWGVF-----DEYNDEKFLYLSNGRIQAVRCASG--ITGTNVVKKCGQSCYKTKR- 209
DB 1 MKWGFILFEGMTLLSGSSYGAD--YIVDNTTYLQVETPGVGQFTL-----NGTEYILED 54

QY 210 -CFNKVVTGLYKGCCEFLVLOSROTEKAS--IMFAOHVDSIVEFCTEQHNHKEAPNKQK 266
DB 55 GDTFTFLEGIYDT--VNMVNRITSLVASCCTVOLNGVIDAIAPSSITGSVNGCTVNLKRG 113

QY 267 CNLRST-----WEVIRDSDFKTTPTMTTQPPNPTFSLQI--GORVCLVLBKS 314
DB 114 CNLSSNITCTLNTGNSRVRNDRILNPASAVMVSGDNVTITSSYINGSAHGIVVNGR 173

QY 315 GSVATGNRLNRLNQAGQLF-----LLQ--TVELGSWGVMTFDSAAHVQSSELIQINS- 365
DB 174 NARIQGNITNSGSGILIHGDCLAEGNTIYLSGDDGVTSDSASGIIRNSISFNSGDG 233

QY 366 ---SDRTLAKRLPAAAGGTST--CSGLRSATFVIRKKYPTDGESEIVLLTDGEDNTISGC 421
DB 234 IRSGNSVITNTVLRNNGTGIYSSGKNATINSNTVKYSGDDG---IYVAGNGSTVQGC 290

QY 422 FNEVKQSGAIHHTVALGPSAAQLEELSKMTGGLQTVASQVONGLIDAFGALSSGNGA 481
DB 291 YVONNTKNGI-----KITSGCTVSSSYTYNG---ENGIYSTGDN 329

QY 482 VSORSIOLESKGLTLQNSQWNGTVIV--DSTVGKDTLFLITWTTPQQLLLMDPSG 536
DB 330 ---SFRYLDSSFNARNGIYSSGKNASFFYITSASYNGENGILSTGPSASMQIIL--DVTS 384

QY 537 QKOGGFVVVKNMAYLOIPGIAKVTGWKYSLOASSOTLTLVTSRASNATLPPIT----- 592
DB 385 NSQNG--VSSRGDNAYIFWEVKKNG--MNGIHSNGKLYLSVKNATNLTSLGINSSTGI 440

QY 593 -----VTSKNTKDTSKFPS---PLVYAN-----IROGASPLRASVT----- 627
DB 441 NARILDVTNLRNGIYSSGYNTTIAYFEAVNNGRNGLITSGRSYIVEGNTTSNRQNG 500

QY 628 -----ALIESVNGKVTLELLONGAGADATKDDGVYSRYFTY-----DTNGRYSV- 673
DB 501 VRVNGPDAIVRSNA-----TDNTASGIA-----IYGNAILIYNCTSMRNTDGVYGIA 548

QY 674 -----KVRALGGVNAARRRVIPO 691
DB 549 SFRMILSRTSANRNCVNAKGTATIEQ 575

RESULT 14
C41859
IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae
C:Species: Haemophilus influenzae
A:Variety: strain HK613
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C:Accession: C41859
J. Bacteriol. 174, 2913-2921, 1992
A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty
A:Reference number: A41859; MUID:92234949
A:Accession: C41859
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1849 <POU>
A:Experimental source: strain HK613
A:Note: sequence extracted from NCBI backbone (NCBIP:97285)
C:Superfamily: IgA-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

C>Date: 23-Sep-1997 #sequence_revision 23-Sep-1997 #text_change 20-Jun-2000
C:Accession: JC5375; PC4485
R:Nakatani, T.; Suzuki, Y.; Yamamoto, T.; Sinohara, H.
J. Biochem. 122, 71-82, 1997
A:Title: Molecular cloning and sequencing of cDNAs encoding three heavy-chain precursors
in inhibitor heavy chain family.
A:Reference number: JC5574; MUID:97420688
A:Accession: JC5575
A:Molecule type: mRNA
A:Residues: 1-946 <NA>
A:Cross-references: DDBJ:D89286; MID:g1694689; PIDN:BAA13939.1; PID:g1694690
A:Experimental source: liver
A:Accession: PC4485
A:Molecule type: protein
A:Residues: 55-64;140-146;151-156;424-447;500-528;577-605 <NA>
C:Comment: In the plasma three inter-alpha-trypsin inhibitor heavy chains 1, 2 and 3 were
that the complexes play important role for pancreatic cancer.
F:Superfamily: inter-alpha-trypsin inhibitor complex component II
F:261-264,717-916/Disulfide bonds: #status predicted

Query Match 2.88; Score 135.5; DB 2; Length 946;
Best Local Similarity 20.1%; Pred. No. 1;
Matches 203; Conservative 124; Mismatches 334; Indels 347; Gaps 54;

QY 20 LSNLIQ--LNNNGYEGIVVAIDPNVPEDE-----LLIQIKDMVTQASLY-L 64
DB 85 MATTIQSKLVNNSPOSONVVEDVQIPKGFISNFTMTVNGITFTSTIREKTGVALYSQ 144
QY 65 FEATGKRFEKNAVAILPETWTKADYVRPKLETYNADVLVAESTPPG---NDEPYTEQ 121
DB 145 ARAGK-----TAGWVRSTLDMENFTEV--NIPGAKVQFELHYQE 185
QY 122 M--GNCGEGERIHLTDFDTAGKLAIEYGP-----QGRAFVHEWAHLRWGVDFDYNDEK 174
DB 186 MKWRKLGSEYEHKIHLPQGLRA--KHLEVNVMIVELQGMRELH-----VPDTF----- 230
QY 175 FYLSNGRIQAVR-CSAGITNVVKK---COGGSCYTKRCTFNKVTGLYEKGCEFFVLSR 230
DB 231 -----EGHFQGVPLSKGKSHSVFRTVAQQRKC--PNCYTYTAVDG----- 271
QY 231 QTEKASIMFAQHVDSEIVEFCTEQHNHKEAPNKQKCNLRSTVEIRDSEDFKKTTPMTT 290
DB 272 -----ELVMVVDVRE-----EKVGELEVFNCYFVHEF 299
QY 291 OPPNPTSLIQICQIRIVCLVLDKSGSMATGNRLNRLNOAGOLF--LQTVELGSGWGMVT 348
DB 300 APEN-----LDPIPKNLFVIDVSGM-WGIKMKQTVEAMKTLDDLRTEDOFS---VVD 350
QY 349 FDSAAHV-QSELQINSGSDRDTLAKRL--PAAASGCTSCSLRSFTVIRK----- 399
DB 351 FNNHVRTWRNDLVSATKTQITD--AKRYIEKIOPSGGTNINEALLRAIFILNEASNLGML 408
QY 400 YPTDGSSEIVLLTDCEDNTISGCFNEVKQSGAIHTVALGPSAAQELSELKMTGGLQTYA 459
DB 409 NPDSVSLVLVSDG-DPTV---GELK-----LSK-----IQKNV 438
QY 460 SDQVONNGLIDAFG-----ALSSGNGAVSQRISQLESKGLFLQNSQWNGMTVIV 508
DB 439 KQNIQDNISLFLSGIDGVDYDFLKLRLSNENRGIAQRI--YGNRDTSSQLKKFYNQ---V 493
QY 509 DSTVGKDTLFLITWTTPQILLWDPGKGGFVVDKNTKMYALQIPGIKVGTYKYSYL 568
DB 494 STPLLRNVQF-----NYPQASVTD-----VTQNSFHNFGGSEIIVVAGKYDPK 537
QY 569 QASQTLTAVTSRASNATLPPIT-----VTSKTNKDTSKPSPPLVWYVANIQR----- 616
DB 538 LAEQVSI-ITATSNTELVELTSLQMDLDFLSKKKHADPNFTKKLWYLTINQLLAER 596
QY 617 GASPILRASVTALIESVNGTV-----TLELLDNGAGADATKDDGVYSRYFT 663
DB 597 SLAP-----TAAIKRKITITLQMSLDHHIVTPTLTAMVINEAGDERMLADSP----- 644

QY 664 TYDTNGYSYKVRALGCVNARRRVIPOQ-----SGALY-----IPGW----- 701
DB 645 -----PDHSCCSCGALYGTGKVASASIPSWASPSPTVM 678
QY 702 ---IENDEIQWNPPEINKDDVOH-----KOVCFSTSGGSGFVSDVPNAPIP 748
DB 679 AMLAVGANRLSETPPHVIRVENDPHFIYLPKSQKNICFNIDSE----- 723
QY 749 DLFPFGQITDLKAEIHGSLINLTWTAPGDDYDHGTAKHYI-----IRISTS 795
DB 724 ---PGKILSLVSDPESGILVN--COLIGAKKAENGKLRTYFGKLCGFVFKEDMKIEISTE 778
QY 796 ILDLRDKFNESLQVNTTALIPKEA---NSEEVFLF--KPNITFENGTDLFIAQVADKV 850
DB 779 NITL---INGS---STTSLFWSDTAHLGNQRVLISVKKGKSVTLTLNKKEMFFSV-LLHHV 831
QY 851 DLKSEISNIARVSLFIPQTPPETPSPOETSAPC---PNIHINSTIPG 895
DB 832 WKKHPV-NVDFLGIYLP--TNKFSFSAHGLLGQFMKPNHIFNERPG 877

Search completed: March 30, 2002, 04:19:20
Job time: 3418 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 04:14:47 ; Search time 63.66 Seconds
(without alignments)
323.092 Million cell updates/sec

Title: US-09-049-696-41
Perfect score: 4759
Sequence: 1 MGPFKSVFLILHLLEGAL.....GIHLKIMKMWIGELQLSLA 914

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1203	25.3	228	1	US-08-469-667-9
2	1203	25.3	228	5	PCT-US95-07289-9
3	140.5	3.0	1529	2	US-08-728-470-10
4	140.5	3.0	1529	4	US-08-719-641-10
5	138.5	2.9	1600	2	US-08-617-697-10
6	137	2.9	1912	1	US-08-409-985-4
7	137	2.9	1912	3	US-08-685-467-4
8	137	2.9	2353	4	US-09-377-155-33
9	137	2.9	2353	4	US-08-913-942-4
10	130	2.7	1848	4	US-08-296-791-6
11	130	2.7	1848	5	PCT-US95-10661A-6
12	128	2.7	1536	1	US-08-038-682-2
13	128	2.7	1536	1	US-08-302-832-2
14	128	2.7	1536	2	US-08-530-198-2
15	128	2.7	1536	2	US-08-469-880-2
16	128	2.7	1536	2	US-08-728-470-2
17	128	2.7	1536	2	US-08-617-697-2
18	128	2.7	1536	4	US-08-719-641-2
19	127.5	2.7	1702	4	US-08-296-791-5
20	127.5	2.7	1702	5	PCT-US95-10661A-5
21	121	2.5	1541	4	US-08-296-791-3
22	121	2.5	1541	5	PCT-US95-10661A-3
23	118.5	2.5	599	3	US-09-045-632-28
24	118.5	2.5	642	3	US-09-045-632-35
25	118.5	2.5	818	3	US-09-045-632-25
26	118.5	2.5	861	3	US-09-045-632-34
27	118.5	2.5	918	3	US-09-045-632-21

28	118.5	2.5	961	3	US-09-045-632-33	Sequence 33, Appl
29	118.5	2.5	1018	3	US-09-045-632-16	Sequence 16, Appl
30	118.5	2.5	1061	3	US-09-045-632-32	Sequence 32, Appl
31	118.5	2.5	1112	3	US-09-045-632-2	Sequence 2, Appl
32	118.5	2.5	1112	3	US-09-045-632-3	Sequence 3, Appl
33	118.5	2.5	1118	4	US-09-379-523-3	Sequence 3, Appl
34	118	2.5	1464	4	US-08-891-640-2	Sequence 2, Appl
35	117	2.5	1447	4	US-09-041-886-25	Sequence 25, Appl
36	117	2.5	1447	5	PCT-US94-05277-2	Sequence 2, Appl
37	116.5	2.4	983	2	US-08-164-292B-26	Sequence 26, Appl
38	116.5	2.4	983	3	US-08-845-623-26	Sequence 26, Appl
39	116.5	2.4	983	3	US-08-815-927-26	Sequence 26, Appl
40	116.5	2.4	1430	3	US-09-008-172-2	Sequence 2, Appl
41	116.5	2.4	1430	4	US-09-210-361-6	Sequence 6, Appl
42	116	2.4	2123	4	US-08-968-685A-10	Sequence 10, Appl
43	115.5	2.4	498	3	US-09-045-632-30	Sequence 30, Appl
44	115.5	2.4	541	3	US-09-045-632-36	Sequence 36, Appl
45	115.5	2.4	621	3	US-08-604-789B-4	Sequence 4, Appl

ALIGNMENTS

RESULT : 1
US-08-469-667-9
Sequence 9, Application US/08469667
Patent No. 5733748
GENERAL INFORMATION:
APPLICANT: Yu, Guo-Liang
APPLICANT: Rosen, Craig
TITLE OF INVENTION: Colon Specific Genes and Proteins
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
ADDRESSEE: Stewart & Olstein
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: NJ
COUNTRY: USA
ZIP: 07068-1739
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,667
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Feitaro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-435
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 228 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-469-667-9

Query Match 25.3%; Score 1203; DB 1; Length 228;
Best Local Similarity 100.0%; Pred. No. 1.2e-96;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 657 VYSRYFTTYDTNGRSVKYKVALGGVNAARRRVIPOQSGALYIPGWIEDEIQWNPPEI 716
|||||
Db 1 VYSRYFTTYDTNGRSVKYKVALGGVNAARRRVIPOQSGALYIPGWIEDEIQWNPPEI 60

```
QY 717 NKDDVQHVKQCFRTSSGGSFVADVPNAPIPDLFPPGQITDLKAEIHGGSGLNLTWTAP 776
Db 61 NKDDVQHVKQCFRTSSGGSFVADVPNAPIPDLFPPGQITDLKAEIHGGSGLNLTWTAP 120
QY 777 GDDYDHGTAHYIIRISTSLDLRDKFNESLOVNTTALIPKEANSEVFLFKPENITTFEN 836
Db 121 GDDYDHGTAHYIIRISTSLDLRDKFNESLOVNTTALIPKEANSEVFLFKPENITTFEN 180
QY 837 GTDLFIATQAVDKYDLKSEISNIARVSLFIPPTPTPSPDTSAPC 884
Db 181 GTDLFIATQAVDKYDLKSEISNIARVSLFIPPTPTPSPDTSAPC 228

RESULT 2
PCT-US95-07289-9
; Sequence 9, Application PC/TUS9507289
; GENERAL INFORMATION:
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Rosen, Craig
; TITLE OF INVENTION: Colon Specific Genes and Proteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; ADDRESSEE: Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07289
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-265
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-07289-9

Query Match 25.3%; Score 1203; DB 5; Length 228;
Best Local Similarity 100.0%; Pred. No. 1.2e-96;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 VYSRYFTTYDNGRYSVKVRALGVNAARRVIPPQSGALYIPGWITENDEIQNPPRPEI 716
Db 1 VYSRYFTTYDNGRYSVKVRALGVNAARRVIPPQSGALYIPGWITENDEIQNPPRPEI 60
QY 717 NKDDVQHVKQCFRTSSGGSFVADVPNAPIPDLFPPGQITDLKAEIHGGSGLNLTWTAP 776
Db 61 NKDDVQHVKQCFRTSSGGSFVADVPNAPIPDLFPPGQITDLKAEIHGGSGLNLTWTAP 120
QY 777 GDDYDHGTAHYIIRISTSLDLRDKFNESLOVNTTALIPKEANSEVFLFKPENITTFEN 836
Db 121 GDDYDHGTAHYIIRISTSLDLRDKFNESLOVNTTALIPKEANSEVFLFKPENITTFEN 180
QY 837 GTDLFIATQAVDKYDLKSEISNIARVSLFIPPTPTPSPDTSAPC 884
Db 181 GTDLFIATQAVDKYDLKSEISNIARVSLFIPPTPTPSPDTSAPC 228
```

```
RESULT 3
US-08-728-470-10
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattiari, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-728-470-10

Query Match 3.0%; Score 140.5; DB 2; Length 1529;
Best Local Similarity 20.9%; Pred. No. 0.0071;
Matches 142; Conservative 94; Mismatches 279; Indels 165; Gaps 29;

QY 281 DFKTTPMTTPTTSLQIGQRIYCLVLDKSGSNATGNRLNR-----LNOAGQ 331
Db 14 DGNKTTIRNSVNAIINWKFQFNIDQNEQFQESSNSAVFNRTSDQISOLKGILDSNGQ 73
QY 332 LFLQ---TVELGSWGMVTFD-SAAHVQSELQINSGSDROTAKRLPAAASGGTSC 386
Db 74 VLLINPNIITCKDAIINTNGFTASTLDISNENIKARNFTLEQTKDALAEIVNHGL-IT 132
QY 387 SGLRSAFTVIRKKYPTD-----GSEIVLLTDGEDNTISGCFNEVKSGAIHTVALGPS 440
Db 133 VKGDSVNLIGGKVKNEGVISNGSGISLLA-GOKITISDIINP-----TITYIAAPEN 186
QY 441 AQAELEELSKWTGGLOTVASQVONNGLI--DAFGALSSGNGAVSORSIQLESKG-LTLQ 497
Db 187 EAINLGDIFAKGGNINVRRA-TIRNKGKLSADSVSKDSGNIVLSAKEGEAIGGVISAQ 245
```

QY 498 NSQWNGTIV-----DSTVGKDTFLITTTTOPPOI 529
Db 246 NOQAKGKLMITGDKVTLTKTGAVIDLSGKGGETYLGDDRGKNGIQAKKTTLEKGS 305
QY 530 LLDWPSQKOGG-----FVVDKNTKMYLQIPGIAKVGTKYSLQASSQTLTL----- 577
Db 306 TI-NVSGKEGGRVWGDIALDGNIN---AQSIAKTGGF---VETSGHDLSTGDDV 358
QY 578 -----TVTSRASNATLPITVTSTKNTKDTSK---FPSPLVYVYANIROG 617
Db 359 IVDKAWLLDPDDVSTETLTSGRNNNGENOGYTTGDKTESPKGNSISKPTLTNSTLEQ- 417
QY 618 ASPILRASVTALIESVNGKVTLEL-LDNCAGADATKDGCVSYRYTFTYDNTGRYSVKVR 676
Db 418 ---ILRRGSYVNITANNRYVNSSINLSNGSLFLTKRDGVKINGDITSNENGLTIKAG 474
QY 677 ALGGVNAARRRVIPOOSGALYIPGWIENDEIOWNPRPE-----INKDD 720
Db 475 SWVDVH-----KNITLTGTFELNI---VAGDSVAFEREGDKARNATDAQITAGTTITVKNKDD 527
QY 721 VQHK--QVCFPSRTSGGSFVSDVPNAPIDLPFPQGIT---DLKAEIHGGSINLT- 772
Db 528 KQFRFNVSINGTGKGLKFIANO-----NNFTHKFDGELNIGSIVTINQTTK 575
QY 773 ---WTAPGDDYH-----GTAHKY---IIRISTSLDLRDKFNESLOVNTTALIPKE 818
Db 576 DVKYWASKDSYWNVSSLTNTVQKFTFKFVDSGNSQDLRSRRSFAGVHENGIGGKT 635
QY 819 -----ANSEEVFLKPNIT 833
Db 636 NFNIGANAKALFKLPNAAT 655

RESULT 4

US-08-719-641-10
; Sequence 10, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-10

Query Match 3.0%; Score 140.5; DB 4; Length 1529;
Best Local Similarity 20.9%; Pred. No. 0.0071;
Matches 142; Conservative 94; Mismatches 279; Indels 165; Gaps 29;
QY 231 DFKKTTTTPMTTQPPNPTFSLIQORIVCLVLDKSGMATGNRLNR-----LNOAQG 331
Db 14 DGNKTTIRNSVNAIINNKOFNIDONEMEQLQESSNSAVFNRYTSQISOLKILDSNGQ 73
QY 332 LFLQ---TVELGSWGMVTPD-SAAHVOSELIQNSGSDRTTLAKRPLPAAASGGTSIC 386
Db 74 VFLINPNGITIGDAIINTNGFTASTLDISNENIKARNFTLEQTKDALKAEIVNHGL-IT 132
QY 387 SGLRSFTVIRKYPD-----GSEIVLLTDCEDNTISCCFNEVKOSGAIHTVALGPS 440
Db 133 VGDGSVNLGGKVGKNEGIVSNGGSISLLA-GQKITISDIINP-----TITTSIAAPEN 186
QY 441 AAQELBELSKMTGGLQTYASDQVQNNGLI--DAFGALSSGNGAVSORSIOLESKG-LTLQ 497
Db 187 EAINLGDIFAKGGINVRAA-TIRNKGKSADSVSKDSGNIVLSAKEGEAIGGVISAQ 245
QY 498 NSQWNGTIV-----DSTVGKDTFLITTTTOPPOI 529
Db 246 NOQAKGKLMITGDKVTLTKTGAVIDLSGKGGETYLGDDRGKNGIQAKKTTLEKGS 305
QY 530 LLDWPSQKOGG-----FVVDKNTKMYLQIPGIAKVGTKYSLQASSQTLTL----- 577
Db 306 TI-NVSGKEGGRVWGDIALDGNIN---AQSIAKTGGF---VETSGHDLSTGDDV 358
QY 578 -----TVTSRASNATLPITVTSTKNTKDTSK---FPSPLVYVYANIROG 617
Db 359 IVDKAWLLDPDDVSTETLTSGRNNNGENOGYTTGDKTESPKGNSISKPTLTNSTLEQ- 417
QY 618 ASPILRASVTALIESVNGKVTLEL-LDNCAGADATKDGCVSYRYTFTYDNTGRYSVKVR 676
Db 418 ---ILRRGSYVNITANNRYVNSSINLSNGSLFLTKRDGVKINGDITSNENGLTIKAG 474
QY 677 ALGGVNAARRRVIPOOSGALYIPGWIENDEIOWNPRPE-----INKDD 720
Db 475 SWVDVH-----KNITLTGTFELNI---VAGDSVAFEREGDKARNATDAQITAGTTITVKNKDD 527
QY 721 VQHK--QVCFPSRTSGGSFVSDVPNAPIDLPFPQGIT---DLKAEIHGGSINLT- 772
Db 528 KQFRFNVSINGTGKGLKFIANO-----NNFTHKFDGELNIGSIVTINQTTK 575
QY 773 ---WTAPGDDYH-----GTAHKY---IIRISTSLDLRDKFNESLOVNTTALIPKE 818
Db 576 DVKYWASKDSYWNVSSLTNTVQKFTFKFVDSGNSQDLRSRRSFAGVHENGIGGKT 635
QY 819 -----ANSEEVFLKPNIT 833
Db 636 NFNIGANAKALFKLPNAAT 655

RESULT 5

US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/08/617,697

APPLICATION NUMBER: 01-APR-1996

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832

FILING DATE: 05-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Berkstreser, Jerry W

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-557

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1600 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-617-697-10

Query Match 2.9%; Score 138.5; DB 2; Length 1600;
Best Local Similarity 20.9%; Pred. No. 0.011;
Matches 142; Conservative 94; Mismatches 279; Indels 165; Gaps 29;

Qy 281 DFKKTTPTTQPPNFTSLQIQCRIVCLVLDKSGMATGNRLNR-----LNOAQQ 331
Db 86 DGNKTTIRNSVNAIINKQFNIDQNEQFQESSAVFNRTSDQISQLGILDSNQ 145
Qy 332 LFLQ----TVELSGWGMVTFD--SAAHVQSELIQINSQSDRTDLAKRLPAAASGTSTC 386
Db 146 VFLNPNGITIGKDAIINTNGFTASTLDSNENIKARNFTLEQTKALAEIVNHGL-IT 204
Qy 387 SGLRSFAFTVRKKYPTD-----GSEIVLLTDGEDNTISCFNEVKQSGAIHTVALGPS 440
Db 205 VGKDSVNLIGKVKNEGVSIVNGSGSISLLA-GOKITISDIINP-----TITYSIAAPEN 258
Qy 441 AAQLEELSKMTGQLOIYASQVQVONGLI--DAFGALSSGNGAVSORSIOLESKG-LTLQ 497
Db 259 EAINIGDIFAKGNINVRRA--TIRNKKLSADSVSKDKSNIVLSAKEGEAIGGVISAQ 317
Qy 498 NSQWMNGTVIV-----DSTVGKOTFLFITWTQPPQI 529
Db 318 NQAAKGLIMITGDKVTLTKCAVIDLSGEGGETYLGGERGEGKNGIQIAKKTILEKGS 377
Qy 530 LWDPSGQKQG-----FVVDKNTKMAVLOIPGIAKVTWKYSIQASSQITLL---- 577
Db 378 TI-NVSGKEKGRATVWGDIADIGNIN---AQGSDIAKTGGF---VETSGHLSIGDDV 430
Qy 578 -----TIVTGRASNATLPPITVTSKNTKDTSK---FPSPLVVYANIRQC 617
Db 431 IVDAKEWLLDPDDVSEIFLTSCRNTNGENQYTTGDKTKEPKGNSISKPTLTNSTLEQ- 489
Qy 618 ASPILRASVTALIESVNGKTVTLEL-LDNGAGADATKDDGVYSRYFTTYDNGRYSVKVR 676

Db 490 ---ILRGSYVNITANNRIYVNSSINLSNGSLTLHTKRDGVKINGDITSNENGLTIKAG 546
Qy 677 ALGGVNAARRRVIPOQSGALYIPGWIEQIWNPPRPE-----INRDD 720
Db 547 SWDVH---KNITLTGTGLNI---VAGDSVAFEREKGDKARNATDAQITAGTITVAKDD 599
Qy 721 VOHK--QVCFSTSSGGSFVADVPNADIPDLFPPGQIT---DLKAEIHGSSLINLT--- 772
Db 600 KQFRENNSVSLNGTGRGLKFIANO-----NNFTHKFDGEINISGLVITNQTTKK 647
Qy 773 ---WTAPGDDYDH-----GTAHKY-----IIRISTSLDLDRKFNESLOVNTTALIPKE 818
Db 648 DVKYNNASKDSTWNVSSLTNTVQKFTFKFVDSGSGNODLRSRRSPAGVHENGIGCKT 707
Qy 819 -----ANSEEVFLFKPENIT 833
Db 708 NENIGANAKALEFKLPNAAT 727
RESULT 6
US-08-409-995-4
Sequence 4, Application US/08409995
Patent No. 5646259
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen I.
APPLICANT: St. Geme III, Joseph W.
TITLE OF INVENTION: Haemophilus Adhesion Proteins
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/409,995
FILING DATE: 24-MAR-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053/RET
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1912 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: unknown
US-08-409-995-4

Query Match 2.9%; Score 137; DB 1; Length 1912;
Best Local Similarity 19.6%; Pred. No. 0.021;
Matches 159; Conservative 103; Mismatches 311; Indels 240; Gaps 39;

Qy 84 TWKTRADYVRPKLEYTKNADVLVAESTPPGNDPEYTEQMCGCKGERIHUTPPFIAGKK 143
Db 428 SWKAKAE-----AD-----TDGALEGISDKQEVKAGE-----TVTFKAGKN 463
Qy 144 LAEYGPQGRAVHE-----WAHLRWGVDFEYNNDEKFFYLS-NGRIQAVRCSAGITGTN 195
Db 464 L-KVKODGANFTYSIQDALTGITSLTGLGTGTTGNDKATVINKDGLTTPAGNGGTGTN 522

QY 196 VVKKCGGSCYTKRCFTNFKVTGL--YEKGEFVLQSRQTEKASTMPAQHVDSIVERCTEQ 253
Db 523 TISVTGDIKAGNKATNVASGLRAYDDANFDVNLNSATD-----LNRHVEDAYKGLL-- 575
QY 254 NHKKEAPNKQKCNLRSTWEVIRDSDEKFTTPMTTPPNTFSLLIQIGQIVCLVLDK 313
Db 576 NLNEKNANKO-----PLVTDSTAATVGDJL---RKLGVVYSTK 609
QY 314 SGSMATGNRLNRUNQAGLFLLOTVELGWSVGMVTFDSAAHVQSELIQINSQ----- 365
Db 610 NG---TKESNOVKQADEV-----LFTGAGAATVTSKSENGKHTITVSVAETKADCGL 659
QY 366 -SDRDLAKRLP-----AAASGTSICSLRSFTVIRKYPDTSGLVLLTIGEDN 416
Db 660 EKGDITIKLVNDQNTDNLVTGNGTAVTKG---GFETV-KTGATDADR-----GKVT 709
QY 417 TISGCFNEVKQSGAIHTHVALGPSAAQAELELSKMTGGLOTYASDOVQNNGLIDAFGALS 476
Db 710 VKDATANDADKKVATKVDKATVATNSAATFVKTNLTTSID---EDNPTDNGKDD---ALK 763
QY 477 SGN-----GAVSORSIOLESKGLTL---QNSQWNGTVIVDSVTKDFTLITWTTPPQI 529
Db 764 AGDTLTFKAGKNLKVDRDGNITFDLAKNLEVAKVSDTLTIGGNT---PTGGTTATPKV 821
QY 530 -----LLWDPGSGQKGGFV---VDKNTKMAVLOI 555
Db 822 NITSTADGLNFAKETADASGSKVYLKGIATTLTETPSAGAKSSHVDLVNVDATKKSNAASI 881
QY 556 PGIAKVGMTKYSLOASSQTLTLTSTRASNAT-----LPPITVTSKNTKDTKFPSPPLV 610
Db 882 EDVLRAG---WNIOGNGNVDYVATVDTVNFDTDDSTGTTVTVTQKADGK----- 929
QY 611 YANIROGASPILRASVTALIESVNGKTVT--LELLDNGAGADATKDDG-----VYSRYF 662
Db 930 -ADVKTGAK-----TSVIKDHNGKLFCTGKLDKDNANGATVSEDDGKDTGTGLVTAK-- 979
QY 663 TTYDTNGRYSVKVRALG-----GVNAARRRVIPOQSGALYIPGWIEDEIOWNPPRP 714
Db 980 TVIDAVKNSGWRVTGEGATAETCATAVNAGNAETVTSVGNFKNG-----NATTA 1030

RESULT 7

US-08-685-467-4
; Sequence 4, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1912 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
POPOLOGY: unknown
MOLECULE TYPE: protein
US-08-685-467-4

Query Match 2.9%; Score 137; DB 3; Length 1912;
Best local Similarity 19.6%; Pred. No. 0.021;
Matches 159; Conservative 103; Mismatches 311; Indels 240; Gaps 39;

QY 384 TWTKADYVRPKLETYNKADVLVAESTPPGNDPEPYEQMNGCCEKGERIHLTPDFIAGKK 143
Db 428 SWKAKAE-----AD-----TDGALEGISKDOEVKAGE-----TVTFKAGKN 463
QY 144 LAYGPGQGRAFVHE-----WAHLRWGVDEYNNDEKFFVLS-NGRIOAVRCSAGLIGTN 195
Db 464 L-KVKQDGANFTYSLQDALTGLTILGGTTNGNDNAKTIVNKDGLTITPAGNGTGTGN 522
QY 196 VVKKCGGSCYTKRCFTNFKVTGL--YEKGEFVLQSRQTEKASTMPAQHVDSIVERCTEQ 253
Db 523 TISVTGDIKAGNKATNVASGLRAYDDANFDVNLNSATD-----LNRHVEDAYKGLL-- 575
QY 254 NHKKEAPNKQKCNLRSTWEVIRDSDEKFTTPMTTPPNTFSLLIQIGQIVCLVLDK 313
Db 576 NLNEKNANKO-----PLVTDSTAATVGDJL---RKLGVVYSTK 609
QY 314 SGSMATGNRLNRUNQAGLFLLOTVELGWSVGMVTFDSAAHVQSELIQINSQ----- 365
Db 610 NG---TKESNOVKQADEV-----LFTGAGAATVTSKSENGKHTITVSVAETKADCGL 659
QY 366 -SDRDLAKRLP-----AAASGTSICSLRSFTVIRKYPDTSGLVLLTIGEDN 416
Db 660 EKGDITIKLVNDQNTDNLVTGNGTAVTKG---GFETV-KTGATDADR-----GKVT 709
QY 417 TISGCFNEVKQSGAIHTHVALGPSAAQAELELSKMTGGLOTYASDOVQNNGLIDAFGALS 476
Db 710 VKDATANDADKKVATKVDKATVATNSAATFVKTNLTTSID---EDNPTDNGKDD---ALK 763
QY 477 SGN-----GAVSORSIOLESKGLTL---QNSQWNGTVIVDSVTKDFTLITWTTPPQI 529
Db 764 AGDTLTFKAGKNLKVDRDGNITFDLAKNLEVAKVSDTLTIGGNT---PTGGTTATPKV 821
QY 530 -----LLWDPGSGQKGGFV---VDKNTKMAVLOI 555
Db 822 NITSTADGLNFAKETADASGSKVYLKGIATTLTETPSAGAKSSHVDLVNVDATKKSNAASI 881
QY 556 PGIAKVGMTKYSLOASSQTLTLTSTRASNAT-----LPPITVTSKNTKDTKFPSPPLV 610
Db 882 EDVLRAG---WNIOGNGNVDYVATVDTVNFDTDDSTGTTVTVTQKADGK----- 929
QY 611 YANIROGASPILRASVTALIESVNGKTVT--LELLDNGAGADATKDDG-----VYSRYF 662
Db 930 -ADVKTGAK-----TSVIKDHNGKLFCTGKLDKDNANGATVSEDDGKDTGTGLVTAK-- 979
QY 663 TTYDTNGRYSVKVRALG-----GVNAARRRVIPOQSGALYIPGWIEDEIOWNPPRP 714
Db 980 TVIDAVKNSGWRVTGEGATAETCATAVNAGNAETVTSVGNFKNG-----NATTA 1030

QY 715 EINKD-----DV-----QHKQVCFRST-----SGSFV-----ASDVPNAPIPDL 750
Db 1031 TVSKDNGNINVKYDVNVGDLKIGDKKIVADTTTLTVTGKVSVPAGANSVNN-----1084
QY 751 FPPGQITDLKAEHGGSLNLTWTAPGDDYDHG 783
Db 1085 ----NKKLVNAEGLATALNNLSWTAKADKYADG 1113
RESULT 8
US-09-377-155-33
; Sequence 33, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAR, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; CURRENT FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-377-155-33

Query Match 2.9%; Score 137; DB 4; Length 2353;
Best Local Similarity 19.6%; Pred. No. 0.029;
Matches 159; Conservative 103; Mismatches 311; Indels 240; Gaps 39;
QY 84 TWTKADYVRPKLETYNADVLVAESTPFGNDPEYTEQMGNGCKEGERHLPDFTAGKK 143
Db 429 SWKAAE-----AD-----TDGALEGISDKQEVKAGE-----TWTFKAGKN 464
QY 144 LAEYGPQGRFVHR-----WAHLRWGVDFEYNNDEKFLS-NGRIQAVRCSAGITGN 195
Db 465 L-KVKQDGANFTSLQDALTLGLTSITLGGTTNGNDKATVINKDKGLTTPAGNGGTTGN 523
QY 196 VVKCKGSGCVTKRCTENKVTGL--YKGCFFVLQSRQTEKASIMFAQHVDSIVEFCTEQ 253
Db 524 TISVTKDGIRKAGNKAITNVASGLRAYDDANFDVILNSATD-----LNRHVEDAYKGLL-- 576
QY 254 NNNKEAPNKQKONLSTWEVIRDSDFKKTTPMTQPPNPTFSLQIGQRIVCLVLDK 313
Db 577 NLNEKNANKQ-----PLVTDSTAATVGDG---RKLGVVWVSTK 610
QY 314 SGSMATGNRLNRLNAQGLFLQIVELGSMVGNVTFPSAAHVQSELIQINSG----- 365
Db 611 NG---TREESNOVROADEV-----LFTGAGAAFTVSKSENGKHITVSVAETKADCGL 660
QY 366 -SDRDTLAKRLP-----AAASCGTSICSLRSFAFTVIRKKYPTDGSSEIVLLTQGEDN 416
Db 661 EKDGDTIKLVQNDQNTDNLVTGNGNGTAVTKG---GFETV-KTGATDADR-----GKVT 710
QY 417 TISGCFNEVKSGALIHFTVALGPSAAQLELSKMTGLQTYASQDVQNNGLIDAFGLS 476
Db 711 VKDATANDADKKVATVKDVAATNSAATFVKTEMLTTSID---EDNPTDNGKDD---ALK 764
QY 477 SGN-----GAVSORSTQLESKGLTL-----QNSQMMNGTIVDSVTKDTFLITWTTPQPI 529
Db 765 AGDTLTFKAGNKLKVKRDKGNTTFLAKNLEVKTAKVSDTLTIGGNT--PTGGTTATPKV 822
QY 530 -----LLWDPSCOKGGFV---VDKNTKWAYLQI 555

Db 823 NITSTADGLNFAKETADASGSKNVYLKGIATTLTEPSAGAKSSHVDLNVDATKKSNAASI 882
QY 556 PGIAKVGTKYKSIQASSQTLTLTTSRSNAT-----LPIITVTSKTKNKTSGKPSPLV 610
Db 883 EDVLRAG---WNIQGNNDVVAIVDTVNFDDSTGTTTIVTQKADGKG-----930
QY 611 YANIROGASPIILRASVTALIESVNGKTVT-LELLDNGAGADATKDDG-----VYSRYF 662
Db 931 -ADVKGIGAK-----TSVIKDHNGKLETKGDKLDKANNAGATVSEDDGDKDTGTGLVTAK-- 980
QY 663 TTYDTNGRYSVKVRALG-----GVNAARRVPOQSGALYIPGWIEDEIOWNPPRP 714
Db 981 TVIDAVNKGSRVTGEGATAETCATAVNAGNAETVTSIGTSVNFKN-----NATTA 1031
QY 715 EINKD-----DV-----QHKQVCFRST-----SGSFV-----ASDVPNAPIPDL 750
Db 1032 TVSKDNGNINVKYDVNVGDLKIGDKKIVADTTTLTVTGKVSVPAGANSVNN-----1085
QY 751 FPPGQITDLKAEHGGSLNLTWTAPGDDYDHG 783
Db 1086 ----NKKLVNAEGLATALNNLSWTAKADKYADG 1114

RESULT 9
US-08-913-942-4
; Sequence 4, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohnbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/4031
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Vance, Dolly A.
; REGISTRATION NUMBER: 39,054
; REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2353 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-913-942-4

Query Match 2.9%; Score 137; DB 4; Length 2353;
Best Local Similarity 19.6%; Pred. No. 0.029;

Matches 159; Conservative 103; Mismatches 311; Indels 240; Gaps 39;

```
QY 84 TWTKKADYVRPKLETYNADVLVAESTPPGNDPEYTEQMGNCGEKGERIHLTPDFIAGKK 143
   ||| |||
Db 429 SWKAAE-----AD-----TDGALEIGSDQEVKAGE-----TVTFKAGN 464
QY 144 LAEYGPQGRAVHE-----WAHLRWGVDFEYNNDEKEYLS-NGRIQAVRCSAGITGN 195
   ||| |||
Db 465 L-KVKQDGANFTYSLQDALTGLSITLGGTTNGNDGAKTVINKDGLTITPAGNGGTGN 523
QY 196 VVKCKCGGSCYTKRCTFNKVTL--YEKGCEFFVLOSROTEKASIMFAQHVDSIVEFECTEQ 253
   ||| |||
Db 524 TISVTRDKIGAKNAITNVASGLRAYDDANFDVNLNSATD-----LNRHVEDAYKGL-- 576
QY 254 NHNKEAPNKNOKNCLNRSTWEVRDSEDFKKTTPMTTPPPNPTFSLLQIQGRIVCLVLK 313
   ||| |||
Db 577 NLNEKNANKO-----PLVTDSTAATVGDL-----RKLGVVSVTK 610
QY 314 SSGMATGNRLNRLNQAGOLFLLQTVELGSMVGMVTFDSSAAHVQSELIQINSG----- 365
Db 611 NG-----TKESNQVQADEV-----LFTGAGAAVTTSKSENGKHTITVSAETKADCGL 660
QY 366 -SDRDLAKRLP-----AAASGTSICSLRSFAFTVIRKPYTDGSEIVLLITDGEDN 416
   ||| |||
Db 661 EKGDITIKLVQNDQNTDNVLTVGNGTAVTKG--GFEIV-KTGATDADR-----GKVT 710
QY 417 TISGCEVNEKQSGAIITHVALGSAQAQELBELSKMTGGLOTYASDQVQNNGLIDAFGLS 476
   ||| |||
Db 711 VKDATANDADKKVATVKDVATAINSAATEFVKTENLTASID--EDNPTONGKDD--ALK 764
QY 477 SGN-----GAVSORSIOLESKGLTL--ONSQWNGTVIVDSVYGVKDTLFLITWTQPPQI 529
   ||| |||
Db 765 AGDTLTFKAGKNLKVRRCKNITFDLAKNLEVKTAKVSDTLTIGGNT--PTGTTTATPKV 822
QY 530 -----LWMDPSGQKQGGFV--VDKNFKMAYLQI 555
   ||| |||
Db 823 NITSTADGLNFAKETADASGKNVYLKGIATLTLEPSAGAKSSHVDLNVDAATKKSNAASI 882
QY 556 PGIAKVGWTKYSLQASSQTLTLTVTSRASNT-----LPITVTSKNTKDTSKFPSPPLV 610
   ||| |||
Db 883 EDVLRAQ--WNIOGNGNVDYVATYDTVNTFTDSDTGTGTTVTVTQKADGK----- 930
QY 611 YANIROGASPILRASVTALIESVNGKTVT-LELLDNGAGADATKDDG-----VYSRYF 662
   ||| |||
Db 931 -ADVKKIAGK-----TSVIRKDHNGKLFYCKDLKDNANGATVSEDDGKDTGTLVTA-- 980
QY 663 TTYDTNGRYSVKVRALG-----GVNAARRRVIPQSGALYIPQWIEIENDEIQWPPRP 714
   ||| |||
Db 981 TVIDAVNKSQWRYTGEATAETGATAVNAGNAETVTSQTSVNFKN-----NATTA 1031
QY 715 EINKD-----DV-----QHKQVCFSRFS--SGGSFV-----ASDVPNAPIPL 750
   ||| |||
Db 1032 TVSKDNGNINVKYDVNVGDLKIGDDKIVADPTTLTVTGGKVSVPAGANSVNN----- 1085
QY 751 FPGQITDLKAEITHGSLINLTWAPGDDYDHG 783
   ||| |||
Db 1086 -----NKKLVNAEGLATALNLSWTAKADKIADG 1114
```

RESULT 10

```
US-08-296-791-6
; Sequence 6, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
```

```
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Irecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RPT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
FLEX: 910 277299
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-296-791-6
```

Query: Match 2.7%; Score 130; DB 4; Length 1848;

Best Local Similarity 21.2%; Pred. No. 0.08;

Matches 203; Conservative 102; Mismatches 299; Indels 352; Gaps 53;

```
QY 109 STPPG--NDPEYTEQ--MGNCG-----EKGERIHL-TPDFIAGKLAEYGPQGRAFY 155
   ||| |||
Db 278 SDPKGILSDPLTNYAVLGDGSPFLFVYDREKQKWLFLGSDYFWAGYNNKSW-QEWNIIYK 336
QY 156 HEWAHLRWGVDFEYNNDEKFLYSLNRIQAVRCSAG-ITGTNNVKKCGGSCYTKRCT--F 212
   ||| |||
Db 337 HEFA-----EKIY-----QOVSAGSLIGSN-----TQYTWQATGST 367
QY 213 NKVTGLYEKCEPVLQSRQTEKASIMFAQHVDSIVEFECTEONHNKEAPNKNQKCNLR-- 270
   ||| |||
Db 368 SIITG-----GGEPLSVDLTDGKD-----KPNHGKSIITLKGSTLNLNH 407
QY 271 -----STWEVIRDSSEDFKKTTPMTTPPPNPTFS-LQIQGRIV 307
   ||| |||
Db 408 IDQAGGLFEEDYEVKGTSDSTTWKAGVSVADGKT--VTWKVHNPKYDRLAKICKG-- 463
QY 308 CLVLD---KSGSMATGNRLNRLNQAGOLFLLQ-----TVELGSMVGMVTFDSSAAHVQS 357
   ||| |||
Db 464 TLVVEGKGKNEGLLKVGD-----GTVLKQKADANNKQVAFSQVIGVSGRST----- 510
QY 358 ELIQINSGSDRDTLAKRLPAAASGGTSSICGLRSFAFTVIRKPYTDGSEIV--LLTDGED 415
   ||| |||
Db 511 --LVLN--DDKQVDPNSIYFGRGRLDLNGLSLTDFHIN--IDGAVRVNHNMTSN 564
QY 416 NTISG-----CFNEVKQSGAIITHVALGPSA 441
   ||| |||
Db 565 ITITGESLITNPNTITSYNIEAQDDDDHPLKIRSIPIRYQLYFNQDNRS---YYTLKKGAST 621
QY 442 AQELEELSKMT-----GGLQTYASDQVON-----NGLIDAFGALSNGNAVRSORI 487
   ||| |||
Db 622 RSELPQNGSESNENWLYMGRTSDAAKRNVMNHNINERNMNGFNQYFG--EETKATONGKL 679
QY 438 QLESKGLTLQNSQWNGMTVIV--DSVVGKDTLFLITWTTPQPOILLWDPSGQKQGGFV-- 543
   ||| |||
Db 630 NVTFNKSKSQNRELLTGGTNELNGDLNVEKGLFSLSRPTPHARDIAGISSTRKDPHFTEN 739
QY 544 -----VDKNKTMAYLIQIPGIAKVGWTKYSLQASSQTLTLTVTSRASN-----A 586
   ||| |||
Db 740 NEVVVEDDMLNRNFKATNMNVTGNA-----SLYSGRVNANITSNITASNAQVHIGYK 792
QY 587 TLPPITVTSK-----TNKDTSKP-PSPLVVYANIROGASPIIL-RASVTALIES 632
```


STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

Query Match 2.7% Score 128; DB 1; Length 1536;
Best Local Similarity 18.8%; Pred. No. 0.087;
Matches 179; Conservative 126; Mismatches 331; Indels 318; Gaps 45;

QY 65 FEAT-----GKRYFKNVAIIIPETWKTKADYVRPKLEYKNADVLVAESTPPGNDPEYTE 120
DB 631 PEGTLNISGK-----VNISWVLP---KNESGYDKFKGRTYWNLTSL----- 668

QY 121 QMNCCKEGERIHLTPDF-----IAGKLAIEYGPQGRAF----- 154
DB 669 ---NVSESGE-FNLTIIDSRGSDSAGTLTPQYNLNGISFNKDTTFNVARNARVNDIKAPI 724

QY 155 -VHEWAHLRWGVFD-----EYN-----NDEKFLYLSNGRIQAVRC 187
DB 725 GINKYSSLVNVSFNGNISVSGGSDVDFLLASSNNVQTPGVVINSKYFNVSSTGSSLRFT 784

QY 188 SAGI-TGTNVVKKCGGSCYTRCTFNKVTVGLYKGCFFVLOSRTKASIMF----- 239
DB 785 SGSTKTGFSIEKD-----LTLNATGGNTITLQVEGTDMIGKGIIVAKKNITFECCNITF 838

QY 240 -----AQHVDISIVEFCTE-----QHNKEAPNKN---QKCNLRSTWEV----- 275
DB 839 GSKAVTEIEGNTVNNANVTLLIGSDFDHQKPLTIKKDVIINSNLTAGGNVNIAGN 898

QY 276 --IRDSEDEKTKTPMT-----TQPNPTFSLQIGQIRIVCLVLDKSGSMATGNRL--- 323
DB 899 LTVESNANFKAITNTFNVGGLFDNKGNSNIAKGARFKIDNKNLSITTNSSSYR 958

QY 324 -----NRLNAGOLFLOTVELGSGWGVTFDSAAHVSQELQ-----INSGSDRDTLAKR 374
DB 959 TIISGNTNKGDLNITNE--GS-----DTEMQIGDVSKQEGNLTSSDKINIKQ 1008

QY 375 --LPAAASGGTSCISGLRSFTVIRKK-----YPTDGESEILVTGDE 414
DB 1009 ITIKAGVDGENSDSDATNANLTIKTKELKLTQDLNISGFNKAETAKDGS---LTIGN 1065

QY 415 DNTISG-----CFNEVKQS--CAIHTVVALGPSAAQLEELSKM--TGGLOTVYASDVON 465
DB 1066 TNSADGTTNAKKVTFNQVKDSKISADGHKVTLH-----SKVETSGNNTEDESDN 1115

QY 466 NGLIDAFGLSSGNGAVSQRSIOLESKGLTLQNSQWNETVIVDSVGVKDTLFLITPTQ 525
DB 1116 NA-----GLIDAKNVVNNITSHKAVSISATSGE-----ITTKT- 1151

QY 526 PPOILLWDPGQKGGFVVDKNTKMYLQIPGIAKVGTWYKYSQASSQILTLITVTSRA-- 583
DB 1152 -----GTTINATTG-----NVEITAQTSILGGIESSGSVTLTATGALA 1192

QY 584 -SNATLPITVTSKTKNDTS-----KFPSPVLV 610
DB 1193 VSNISGNTVTVTANSAGLTLTAGTIKGTESVITSSOSGDIGGTISGTVVKTATESLIT 1252

QY 611 YANIR---QGASPTLRASVTALIESVNGKVTLE-----LIDNGAGADATK----- 653
DB 1253 QNSKIKATTCEANVTSATGT-IGGTISGNTVNVNANAGDLTVGNGAEINATEGAATLTT 1311

QY 654 DGVVSRVFTTYDNGRYSKVRA-----LGVNNAARRRVIPOOSGALYIPGWIENDEIQ 708
DB 1312 SSGKLTEASSHITSAGQVNLDAQDSGAVGSINAA--NVTNLNTGTLTT---VKGSNIN 1366

QY 709 WNPPELKNDDVQ-----HQVCFSTRSSGGSFVSDVPNAPIP-DLFPFGQI--- 756
DB 1367 ATSGTLVINAKDAELNGALGNHTVNVNATNANGSSVITATSSRVNITGDLITINGLNI 1426

QY 757 --TDLKAIEHSGSLNLTWTAPG--DDYDHTGAHYIIRISTISILDLDKKNESL 807
DB 1427 SKNGINTVLLKGVKIDVKYIQPGIASVDEVIETAK---RILEKVKDLSDEREAL 1477

RESULT 14
US-08-530-198-2
Sequence 2, Application US/08530198
Patent No. 5869065

GENERAL INFORMATION:

APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,198
FILING DATE: 13-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: JWB-1186
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-530-198-2

Query Match 2.7% Score 128; DB 2; Length 1536;
Best Local Similarity 18.8%; Pred. No. 0.087;
Matches 179; Conservative 126; Mismatches 331; Indels 318; Gaps 45;

QY 65 FEAT-----GKRYFKNVAIIIPETWKTKADYVRPKLEYKNADVLVAESTPPGNDPEYTE 120
DB 631 PEGTLNISGK-----VNISWVLP---KNESGYDKFKGRTYWNLTSL----- 668

QY 121 QMNCCKEGERIHLTPDF-----IAGKLAIEYGPQGRAF----- 154
DB 669 ---NVSESGE-FNLTIIDSRGSDSAGTLTPQYNLNGISFNKDTTFNVARNARVNDIKAPI 724

QY 155 -VHEWAHLRWGVFD-----EYN-----NDEKFLYLSNGRIQAVRC 187
DB 725 GINKYSSLVNVSFNGNISVSGGSDVDFLLASSNNVQTPGVVINSKYFNVSSTGSSLRFT 784

QY 188 SAGI-TGTNVVKKCGGSCYTRCTFNKVTVGLYKGCFFVLOSRTKASIMF----- 239
DB 785 SGSTKTGFSIEKD-----LTLNATGGNTITLQVEGTDMIGKGIIVAKKNITFECCNITF 838

QY 240 -----AQHVDISIVEFCTE-----QHNKEAPNKN---QKCNLRSTWEV----- 275
DB 839 GSKAVTEIEGNTVNNANVTLLIGSDFDHQKPLTIKKDVIINSNLTAGGNVNIAGN 898

QY 276 --IRDSEDEKTKTPMT-----TQPNPTFSLQIGQIRIVCLVLDKSGSMATGNRL--- 323
DB 899 LTVESNANFKAITNTFNVGGLFDNKGNSNIAKGARFKIDNKNLSITTNSSSYR 958

QY 324 -----NRLNAGQGLFLLQTVELGWSWGMVTFDSAHHVQSELIO-----INSGSDRDLAKR 374
Db 959 TIISGNITNKGDLNITNE---GS-----DTEMOIGGVDSQKEGNLTSSDKINITQ 1008
QY 375 --LPAAASGGSICSLRSFAFTVIRKK-----YPTDGSSEIVLLTDGE 414
Db 1009 ITIKAGVDGENSDSDATNANLTIKTELKLTQDLNLSGFKAEITAKDGS---LTIGN 1065
QY 415 DNTISG-----CFNEVKQS--GAIHTVALGPSAAQEELEESKM-TGGLOTYASDOVON 465
Db 1066 TNSADGTNAKKVTFNQVKDSKISADGHKVTLH-----SKVETSGSNNTEDSSDN 1115
QY 466 NGLIDAFGALSSGNGAVSQRSIOLESKGLTLQNSOMWNGTVIVDSVGRDTEFLITWTQ 525
Db 1116 NA-----GLTIDAKNVTNNNITSHKAVSISATSGE-----ITTKT- 1151
QY 526 PPOILLWDPGQKGGVVDKNTKMAYLQIPGIKAKYTKWYSLQASSQTLTLVTTSRA-- 583
Db 1152 -----GTTINATTG-----NVEITAQTGSLGGIESSGVTLTATEGALA 1192
QY 584 -SNATLPPITVTSKTNKDTs-----KFPSPLVV 610
Db 1193 VSNISGNTVTVTANSALITLAGSTIKGTESVTTSSQSGDIGGTISGGTVEVKATESLT 1252
QY 611 YANIR---QGASPIRLASVTALIESVNGKVTYLE-----LLDNGAGADATK----- 653
Db 1253 QNSKIKATGEANVTSAITG-IGGTISGNTVNTVANAGDLTVGNGAEINATEGAATLT 1311
QY 654 DGVVSRVETTYDTNGRYSVKVRA-----LGGVNAARRRVIPOQSGALYIPGWIENDEIQ 708
Db 1312 SSGKLTTEASSHITSAGQVNLQAQDSVAGSINAA--NVTLNTTGLIT--VKGSNIN 1366
QY 709 WNPPEIKNDQV-----HKQVCFRTSSGGSFVADSVNADIP-DLPPPGQI--- 756
Db 1367 ATSGTLVINAKDAELNGAALGNHTVYNATNANGSGSVIATTSRVAITGDLITINGLINII 1426
QY 757 --TDLKAEITHGGLINLTWAPG-DDYDHGTAHKYIIRISTSLDRLDRFNESL 807
Db 1427 SRKNGINTVLLKGKVIDKYIQGSIASVDEVEIAK---RILEKVKYKLSDBEREAL 1477

RESULT 15

US-08-469-880-2
; Sequence 2, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:V9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-2

Query Match 2.7%; Score 128; DB 2; Length 1536;
Best local Similarity 18.8%; Pred. No. 0.087;
Matches 179; Conservative 126; Mismatches 331; Indels 318; Gaps 45;
QY 55 FEAT-----GRFYEKNVAIIPETWTKADYVRPKLETYKNADVLVAESTPPGNDPEYTE 120
Db 631 FEGLTNISGR-----VNISMVLP--KNESGYDKPKGRTYWNLTSL----- 668
QY 121 QMNCCEGKEGRIHLTPDF-----TAGKLAEYGPQGRAF----- 154
Db 669 ---NVESGE-FNLTIDSRGSDSAGTLTPQYNLNGISFNKDTTFNVERNARVNFDIKAPI 724
QY 155 -VHEWAHLRWGVFD-----EYN-----NDEKPYLSNGRIQAVRC 187
Db 725 GINKYSLNATVPNGNISVSGGSDVFTLLASSNSVOTPGVVINSKYFNVSSTSLRFT 784
QY 198 SAGI-TGTNVVKCOGSCYTKRFTFNKVTGLYEKGCEFLVLOSQTAKASIMF----- 239
Db 785 SGSTKYGFSLEKD-----LTLNATGNITLLQVEGDMIGKGVAKKNITFFEGGNITP 838
QY 240 -----AQHVDSIVFEFCTE-----QHNKEAPKNQN---OKCNLRSTWVEV----- 275
Db 839 GSRKAVTEIEGNTVINNNANVTLLIGSDFDNHQKPLTIKKDVIINSNLTAGGNIVNTAGN 898
QY 276 --IRDSDFPKTTPMT-----TQPPNPTFSLQIGORIVCLVLDKSGSMATGNRL--- 323
Db 899 LTVESNANFKAITNFTFNVGGLFDNKGNSNISTAKGGARFKDIDNSKNLSITTNSSSTYR 958
QY 324 -----NRLNAGQGLFLLQTVELGWSWGMVTFDSAHHVQSELIO-----INSGSDRDLAKR 374
Db 959 TIISGNITNKGDLNITNE---GS-----DTEMOIGGVDSQKEGNLTSSDKINITQ 1008
QY 375 --LPAAASGGSICSLRSFAFTVIRKK-----YPTDGSSEIVLLTDGE 414
Db 1009 ITIKAGVDGENSDSDATNANLTIKTELKLTQDLNLSGFKAEITAKDGS---LTIGN 1065
QY 415 DNTISG-----CFNEVKQS--GAIHTVALGPSAAQEELEESKM-TGGLOTYASDOVON 465
Db 1066 TNSADGTNAKKVTFNQVKDSKISADGHKVTLH-----SKVETSGSNNTEDSSDN 1115
QY 466 NGLIDAFGALSSGNGAVSQRSIOLESKGLTLQNSOMWNGTVIVDSVGRDTEFLITWTQ 525
Db 1116 NA-----GLTIDAKNVTNNNITSHKAVSISATSGE-----ITTKT- 1151
QY 526 PPOILLWDPGQKGGVVDKNTKMAYLQIPGIKAKYTKWYSLQASSQTLTLVTTSRA-- 583
Db 1152 -----GTTINATTG-----NVEITAQTGSLGGIESSGVTLTATEGALA 1192
QY 584 -SNATLPPITVTSKTNKDTs-----KFPSPLVV 610
Db 1193 VSNISGNTVTVTANSALITLAGSTIKGTESVTTSSQSGDIGGTISGGTVEVKATESLT 1252

611	QY	YANIR---QGASPILRASVTALJESVNGKVTLE-----LIDNGAGADATK-----	653
1253	Db	QNSKIKATTCEANVTSATGT-IGGTTGNTVNVNANAGDLTVNGGAEINATEGAALTT	1311
654	QY	DGVVSRTFTYDNGRYSKVRA-----LGGVNAARRRVIPOOSGALYIPGWIENDEIQ	708
1312	Db	SSGKLTTEASSHITSAGQVNLSDQGSVAGSINAA--NVTLNTTGTTLTT---VKGSNIN	1366
709	QY	WNPPRPETNKDDVQ-----HKQVCFSRITSSGSGSVASDVPNAPIP-DLFPFGOI---	756
1367	Db	ATSGTILVINAKDAELNGAALGNHTVVNATNANGSGSVIATTSRVNTTGLITINGLNII	1426
757	QY	--TDUKAEIHGSSLNLNLTWTPAG-DDYDHGTAHKYIIRISTSIILDLRDKNESIL	807
1427	Db	SKNGINTVLKGVKIDVYIOPGIASVDEVIEAK---RILEKVKADLSDEFERFAL	1477

Search completed: March 30, 2002, 04:20:38
Job time: 351 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 03:02:02 ; Search time 78.48 Seconds
(without alignments)
862.678 Million cell updates/sec

Title: US-09-049-696-41
Perfect score: 4759
Sequence: 1 MGPPKSSVFILILHLEAL.....GIHLKIMWKIGELQLSLA 914

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4759	100.0	914	20 AAB74824	Human ICACC-1 prot
2	4759	100.0	914	22 AAB74733	Human secreted pro
3	4759	100.0	925	22 AAG75474	Human colon cancer
4	4756	99.9	914	22 AAM24514	C902P predicted am
5	4756	99.9	914	22 AAB73716	Human CLCA1 protei
6	4776	94.1	869	22 AAG75614	Human colon cancer
7	3656.5	76.8	913	20 AAB74822	Murine ICACC-1 pro
8	3656.5	76.8	913	22 AAB73715	Mouse Gob-5 protei
9	2879.5	60.5	919	21 AAY66749	Membrane-bound pro
10	2879.5	60.5	919	22 AAB87560	Human PRO1124, Ho
11	2879.5	60.5	919	22 AAB65272	Human PRO1124 (UNQ

12	2831	59.5	552	22 AAG73854	Human colon cancer
13	1995.5	41.9	943	20 AAY41036	Human lung tumor a
14	1995.5	41.9	943	21 AAB11321	Human lung cancer-
15	1986.5	41.7	943	20 AAB74823	Human ICACC-2 prot
16	1960	41.2	942	21 AAY70459	Human membrane cha
17	1782.5	37.5	791	20 AAY41040	Human lung tumor a
18	1782.5	37.5	791	21 AAB11325	Human lung cancer-
19	1518.5	31.9	742	22 AAB45904	Human tumor-associ
20	1340.5	28.2	592	20 AAY41039	Human lung tumor a
21	1340.5	28.2	592	21 AAB11324	Human lung cancer-
22	1203	25.3	228	18 AAW06548	Human colon specif
23	1203	25.3	228	19 AAW46879	Protein sequence e
24	645.5	13.6	218	21 AAB53310	Human colon cancer
25	645.5	13.6	218	22 AAG73761	Human colon cancer
26	199	4.2	92	20 AAY11789	Human 5' EST seque
27	193	4.1	91	21 AAG03643	Human secreted pro
28	143.5	3.0	1601	18 AAW30292	Non-typeable Haemo
29	143	3.0	2411	21 AAB23860	Haemophilus influe
30	140.5	3.0	1529	14 AAR41732	High molecular wei
31	139.5	2.9	2039	19 AAW56322	Haemophilus paraga
32	137	2.9	2353	17 AAR99393	Haemophilus adhesi
33	135.5	2.8	1152	22 AAG89860	C glutamicum prote
34	135.5	2.8	1152	22 AAB79787	Corynebacterium gl
35	134	2.8	26926	22 AAU05396	Human titin (conne
36	133.5	2.8	1569	22 AAG98842	E. coli growth and
37	131	2.8	498	22 AAB49641	Escherichia coli H
38	131	2.8	1095	21 AAB01835	Haemophilus influe
39	131	2.8	1101	21 AAB01834	Haemophilus influe
40	130.5	2.7	505	21 AAB15522	S. muenchen flagel
41	129.5	2.7	969	21 AAB01827	Haemophilus influe
42	129.5	2.7	975	21 AAB01826	Haemophilus influe
43	129	2.7	498	8 AAP70302	Sequence of flagel
44	128.5	2.7	605	22 AAB49643	Escherichia coli H
45	128	2.7	1095	21 AAB01847	Haemophilus influe

ALIGNMENTS

RESULT 1
AAB74824
ID AAB74824 standard; Protein; 914 AA.
XX
AC AAB74824;
XX
DT 13-JUN-2001 (first entry)
XX
DE Human ICACC-1 protein sequence.
XX
KW ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW Interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW inflammatory bowel disease; autoimmune disease.
XX Homo sapiens.
XX
PN WO9944620-A1.
XX
PD 10-SEP-1999.
XX
PF 03-MAR-1999; 99WO-US04703.
XX
PR 03-MAR-1998; 98US-0076815.
XX
PA (MAGA-) MAGAININ PHARM INC.
XX
PI Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;
XX Nicolaides NC, Zhou Y, Dong Q;
XX WPI; 1999-550979/46.
XX N-PSDB; AAF81927.

Db	601	tskfspplvvyanirgaspilrasvtaliesvngkvtvleldngagadatkdgvysr	660
Qy	661	yfttydntngrysvkvralsgvnaarrvipoqsgalytpgwiendeiqwnpprpeinkdd	720
Db	661	yfttydntngrysvkvralsgvnaarrvipoqsgalytpgwiendeiqwnpprpeinkdd	720
Qy	721	vohkqvcsfrrtssggsfvasdvpnapipdlpppgoitldkkaeiHGSLINTWTAPGDDY	780
Db	721	vqhkqvcfsrtssggsfvasdvpnapipdlpppgoitldkkaeiHGSLINTWTAPGDDY	780
Qy	781	dhgtahkyiiriststldlrdkfneslqvnttallpkkeanseevflfxpenltfengtdl	840
Db	781	dhgtahkyiiriststldlrdkfneslqvnttallpkkeanseevflfxpenltfengtdl	840
Qy	841	ftiaqvkvdlkkselnsrtarvslrptpdpetsapcniHINStipgihlK	900
Db	841	flaigavkvdikkselnsrtarvslrptpdpetsapcniHINStipgihlK	900
Qy	901	imkwkigelqlslA 914	
Db	901	imkwkigelqlslA 914	
RESULT	2		
ID	AAB74733		
XX	AAB74733 standard; Protein; 914 AA.		
AC	AAB74733;		
XX	12-JUN-2001 (first entry)		
DT			
XX	Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.		
DE	Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;		
KW	dermatological; immunosuppressive; antiinflammatory; anti-HIV;		
KW	immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;		
KW	ophthalmological; neuroprotectant; antitropic; anticonvulsant; vaccine;		
KW	antialzheimer; antiparkinsonian; antimicrobial; vulnervary; gene therapy;		
KW	immune disorder; hyperproliferative disorder; cardiovascular disease;		
KW	cancer; angiogenic disorder; neurological disorder; infectious disease;		
KW	wound healing; regeneration; chemotaxis; chromosome 1.		
OS	Homo sapiens.		
XX	WC200112775-A2.		
PN	22-FEB-2001.		
PD	16-AUG-2000; 2000WO-US22325.		
XX	17-AUG-1999; 99US-0149182.		
XX	(HUMA-) HUMAN GENOME SCI INC.		
PA	Rosen CA, Ni J, Florence KA, Fiscella M, Wei P, Baker KP;		
XX	Birze CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;		
PI	WPI; 2001-147550/15.		
XX	N-PSDB; AAF81787.		
DR	Nucleic acids encoding 25 human secreted polypeptides, useful for		
XX	preventing, diagnosing and/or treating e.g. cancers, parkinson's		
PT	disease and diabetic retinopathy -		
PS	Claim 11; Page 459-460; 485pp; English.		
XX	AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733		
CC	to AAB74772. Human secreted proteins can have activities based on the		
CC	tissues and cells they are expressed in. Example of activities include:		
CC	immunomodulatory; antisclerotic; dermatological; immunosuppressive;		
CC	antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;		
CC	vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic;		
CC	anticonvulsant; antialzheimer; antiparkinsonian; antimicrobial; and		

New nucleic acid encoding calcium activated chloride channel, used to identify, e.g. specific modulators for treating atopic allergy -

Claim 11; Fig 4B; 75pp; English.

The present sequence represents the human interleukin 9 (IL-9) induced calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins have anti-allergic, anti-asthmatic, anti-inflammatory and immunomodulatory activities. Compounds (A) that downregulate ICACC are used to alleviate asthma (or more generally atopic allergy), while those (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and inflammatory bowel disease (IBD) (or other autoimmune diseases). Measuring levels of ICACC-1 can be used in the diagnosis of asthma (increased levels) or IBD (reduced levels), also for monitoring treatment of these conditions. The ICACC proteins can be used: (i) to raise specific antibodies (Ab), useful: (a) as immunoassay reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as (A) to disrupt bonding between ICACC and its ligands; and (iii) to identify modulators and binding partners. ICACC polynucleotides can be used to generate transgenic animals or recombinant cells, used to screen for antagonists, also as a source of therapeutic antisense agents or diagnostic probes (for quantifying mRNA expression, e.g. for identification of modulators).

Sequence 914 AA;

Query Match	100.0%;	Score 4759;	DB 20;	Length 914;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 914;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
1	MGPFKSVFTLIHLLEGALSLSLIQNNNGYEGIVAIIDPNVPDETLIQIKDMVTOA	60		
1	mgpfksvftlihllegalslsliqnnngyegivaidpnvpdedtliqikdmvtga	60		
61	SLVLFATGKRFYKFNVAIIPETWTKADYVRPKLETYNADVLVAESTPGNDEPYTE	120		
61	slylfatgkrfykfnvailipetwtkadyvrpkletynadvlvaestpgndepyte	120		
121	QMNGCKGKGRHILTPDIFAGKGLAEYGPQGRFVHWAHLRWGVFDEYNDEKPYLSNG	180		
121	qmngckgkgrihltpdifagkglaeypqgrfwhwahrwgvfdeynndekfyslsg	180		
181	RIQAVRCAGITGNVVKYKCGGSCYTKRCTFNKVTGLYKCEFFVLOSQTKEASIMFA	240		
181	riqavrcagitgnvvkycggscytkrctfnkvtgklykceffvlosqtkeasimfa	240		
241	QHVDSEVFECTEONHNKEAPNKNOKNLRSTWEVIRDSDEKTKTTPMTTQPPNPTFSL	300		
241	qhvdsevfcteqnhnkeapnknoknlrstwevirdsedfkktpmttqppnptfsl	300		
301	QTGORIVCLVLDKSGMATGNRLNRLNOAGQLFLLQTVLGSVGMVTFDSAHHVOSLTI	360		
301	qlgqriyclvldksgmatgnrlnrlnqagqlfllqtvlgsvgmvtfdsaahvgseli	360		
361	QINSGSDROTTLAKRLPAAAGSGTSCISGLRSFAFTVIRKYPDTGSEIVLTLTGDENITISG	420		
361	qinsgsdrottlakrlpaaasggtscisglrsaftvirkypdtgseivlltdgedntisg	420		
421	CFNEVKQSGAIHTHVALGPSAAQLEELSKMTGGLQTVASDVQVONNGLIIDAFALSSGNG	480		
421	cfnevkqsgaihtvalgpsaaqleelskmtgglqtvasdqvonggliidafalssng	480		
481	AVSQRISQLESKGLILONSOMNGVIVDSTVGKDTFLITWTTPPQIILLWDPSPGKQK	540		
481	avsqrisqlskglilqnsqwmngvividstvgkdtflitwtqppqillwdpsgkqg	540		
541	GFVVDKNTKMYLIQIPGIAKVGTWKYSLOASQTLTLTIVTSRASNLPTTIVTSKTNKD	600		
541	gfvdntkmyliqipgiakvgtwkyisqassqtlitlvtvtsrasnatlpttvtvtsktnkd	600		
601	TSKFSPFLVYVYANIRGASPIILRASVTALIESVNGKTVTLELDNGAGADATKDDGVYSR	660		

601	tskfspflvvyanyirgaspillrasvtalliesvngktvtlleldngagadatkdgvysr	660
661	yfttydntgrysvkvralgsgvnaarrvpiqosgalyipgwiendeiqwnpprpeinkdd	720
661	yfttydntgrysvkvralgsgvnaarrvpiqosgalyipgwiendeiqwnpprpeinkdd	720
721	vohkqvcsrstsrgsgsvasdvnpnapidlpfpgqitdlkaethgslinltwtapgddy	780
721	vohkqvcsrstsrgsgsvasdvnpnapidlpfpgqitdlkaethgslinltwtapgddy	780
781	dhgtahkviiristildrkeneslqvntalipkeanseevflfkpenitfengtdl	840
781	dhgtahkviiristildrkeneslqvntalipkeanseevflfkpenitfengtdl	840
841	ftiaqvadkvdlskserniarvslfipqtpptpsetpsdetsapcpnhinstipgihlk	900
841	ftiaqvadkvdlskserniarvslfipqtpptpsetpsdetsapcpnhinstipgihlk	900
901	imkwigeliglsia	914
901	imkwigeliglsia	914
ULT 2		
AAB74733		
AAB74733 standard; Protein; 914 AA.		
AAB74733;		
12-JUN-2001 (first entry)		
Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.		
Human; secreted protein; diagnosis; immunomodulatory; antisclerotic; dermatological; immunosuppressive; antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic; antimicrobial; vaccine; antialzheimer; antiparkinsonian; antidiabetic; cardiovascular disease; immune disorder; hyperproliferative disorder; neurovascular disease; cancer; angiogenic disorder; neurological disorder; infectious disease; wound healing; regeneration; chemotaxis; chromosome 1.		
Homo sapiens.		
WC200112775-A2.		
22-FEB-2001.		
16-AUG-2000; 2000WO-US22325.		
17-AUG-1999; 99US-0149182.		
(HUMA-) HUMAN GENOME SCI INC.		
Rosen CA, Ni J, Florence KA, Fiscella M, Wei P, Baker KP; Birse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR; WPI; 2001-147550/15.		
N-PSDB; AAF81787.		
Nucleic acids encoding 25 human secreted polypeptides, useful for preventing, diagnosing and/or treating e.g. cancers, Parkinson's disease and diabetic retinopathy -		
Claim 11; Page 459-460; 485pp; English.		
AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733 to AAB74772. Human secreted proteins can have activities based on the tissues and cells they are expressed in. Example of activities include: immunomodulatory; antisclerotic; dermatological; immunosuppressive; antinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic; anticonvulsant; antialzheimer; antiparkinsonian; antimicrobial; and		

CC vulnary. Human secreted proteins can be used in gene therapy and
CC vaccine. Human secreted protein nucleotide sequences (NAMI) and proteins
CC (PEPI) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. For example, NAMI
CC and PEPI may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patients genome
CC that affect the activity of proteins by expressing inactive proteins or
CC to supplement the patients own production of polypeptides. Disorders that
CC may be prevented, diagnosed and/or treated include immune disorders,
CC hyperproliferative disorders (e.g. cancers), cardiovascular diseases,
CC angiogenic disorders, neurological disorders, infectious diseases and/or
CC for promoting wound healing, regeneration and/or chemotaxis. AAF81778 to
CC AAF81786 and AAB74732 represent sequences used in the exemplification of
CC the present invention.

XX SQ Sequence 914 AA;

Query Match 100.0%; Score 4759; DB 22; Length 914;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 914; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPFKSSVFILHLLEGALSLSLIQNNNGYEGIVVAIDPNVPEDETLIQIQIKDMVTOA 60
DB 1 mgpfkssvfllhllegalslsliqnnngyegivvaiddpnvpedetliqikdmvtqa 60
QY 61 SLYLFEATGKRFYKNAVILPETWTKADYVRPKLETYKNADVLVAESTPPGNDSPYTE 120
DB 61 slylfeatgkrfyknavilpetwtkadyvrpkletyknadvlvaestppgndepyle 120
QY 121 QMNGCCKGERTHLTPDFTAGKFLAEGPQGFVHWAHLRWGVFDEYNNDKFFVLSNG 180
DB 121 qmngcckgerthltpdftagkflaegpqqgrafvnewahlrwgvfdeyndndekfrylsng 180
QY 181 RTQAVRCSAGITCTNVVKKCCQGCSCYTKRCTFNKNVGLYKGCFFVLQSRQTEKASIMFA 240
DB 181 riqavrcsagitctnvvkkccqgcscytkrctfnknvglkyekgcefvlsqrqtekasimfa 240
QY 241 QHVDSTVEFCTONHNKAPNQNOKCNLRSTWEVTRDSEDFKTKTPMTQPPNPTFSL 300
DB 241 qhvdstvefcteqnhkeapnqknqcnlrstwevtrdsedfkktpmttqppnptfsl 300
QY 301 QIQQRIVCLVLDKSGSMATGNRLNRLNQAGOLFLLQTVLGSWMGVMTFDSAHVOSLI 360
DB 301 qiqqrivclvldksgsmatgnrlnrlnqagqlfllqcvlgswwgmvtfdsaahvgseli 360
QY 361 QINSGSDRDLAKRLPAASAGTSCISGLRSAPTVIRKKYPTDGSFVILLTDGEDNTISG 420
DB 361 qinsgsdrdlakrlpaaasagtsicsglrsaftvirkkypdgsfvlitdgedntisg 420
QY 421 CFNEVKQSGAILHTVALGPSAAQLEELSKMTGCGLOTYASDQVQNNGLIDAFGALSSNG 480
DB 421 cfnevkqsgailhtvalgpsaaqleelskmtgglqtyasdqvgngnllidafgalsng 480
QY 481 AVSORSIOLESKGLTLQNSQWNGTVIVDSTVCKDFLITWTTPQQLLLWDPSSQKQG 540
DB 481 avsqrsioleskgltlqnsqwmngtvivdstvckdflitwtppqqlldwpsqkgg 540
QY 541 GFVVDKNTMAYLIQPIGIAKVGWTKVLSQASSOTLTLVTSTRASNATLPPITVTSKTKD 600
DB 541 gfvdnkntmayliqpigakvgwtkvlsqassotlcltvtstrasnatlppitvtsktnkd 600
QY 601 TSKFSPPLVYANIRGASPIRLASVTALIESVNGKTVILELLDNCAGADATKDDGVYSR 660
DB 601 tskfspplvyanirgaspiirlasvtaliesvngkvtvilelldngagadatkdgvysr 660
QY 661 YFTTYDTNGRYSVKVVALGVNAARRVTPQOSGALYIPGWIENDEIOWNPPRPELNKDD 720
DB 661 yfttydtngrysvkvvalgvnaarrvtpqdsгалyipgwiendeiownprrpeinkdd 720
QY 721 VQHKQVCFSTRSGGSFVASDVPNAPIPDLFPPGOITDLKAEIHGGSLLINTWTAPGDDY 780
DB 721 vqhkvcfstrsggsfvasdvpnapipdlfppgoitdlkaeihggsllintwtapgdyy 780

QY 791 DGHATAKYIIRISTSTILDRLDKFNESLQVNTTALIPKEANSEVFLFKPENITFENGTDL 840
DB 791 dghatakyiiristsildrldkfneslqvnttalipkeanseeflfrkpenitfengtdl 840
QY 841 FTAIOQAVDKVLDKSEISNTARVSLFPPQTPPETSPDETSAFCPNTHINSTIPGTHILK 900
DB 841 ftaioqavdkvldkseisntarvslfppqtppetspdetfapcnpnthinstipgthilk 900
QY 901 IMWKWIGELQLSIA 914
DB 901 imkwkigelqlsia 914

RESULT 3

AAAG75474
ID AAG75474 standard; Protein; 925 AA.

XX AC AAG75474;

XX DT 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:6238.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; chromosome 1.

XX Homo sapiens.

XX WO2000122920-A2.

XX PD 05-APR-2001.

XX PF 28-SEP-2000; 2000WO-US26524.

XX PR 29-SEP-1999; 99US-0157137.

XX PR 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI: 2001-235357/24.

XX N-PSDB; AAH34879.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -

XX Claim 11; Page 7686-7690; 9803pp; English.

XX AAH32943 to AAH37195 and AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing P.
XX Inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAB77789 represent sequences used in the exemplification of the
XX present invention.

XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 925 AA;

AC	AA24514;																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										</
----	----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

AC	AAM24514;				
XX	12-OCT-2001	(first entry)			
DT					
XX	C902P	predicted amino acid sequence.			
DE					
XX	Human;	immunotherapy; diagnosis; colon cancer; colon tumour;			
KW		immunogenic; gene therapy; vaccine; colonic cancer.			
XX					
OS		Homo sapiens.			
XX					
PN	WO200149716-A2.				
XX	12-JUL-2001.				
XX	29-DEC-2000;	2000WO-US35596.			
EF					
XX	30-DEC-1999;	99US-0476296.			
PR	10-JAN-2000;	2000US-0480321.			
PR	15-FEB-2000;	2000US-0504629.			
PR	06-MAR-2000;	2000US-0519444.			
PR	19-MAY-2000;	2000US-0575251.			
PR	29-JUN-2000;	2000US-0609448.			
PR	28-AUG-2000;	2000US-0649811.			
XX					
XX	(CORI-) CORIXA CORP.				
PA					
XX	Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;				
PI	King GE, Wang T, Jiang Y;				
XX					
DR	WPI; 2001-441847/47.				
XX					
XX	Colon tumor associated proteins and nucleic acids useful for the				
PT	prevention, diagnosis and treatment of colonic cancer -				
PT					
XX					
PS	Claim 2; Page 440-443; 472pp; English.				
XX					
CC	The present invention describes colon tumour associated proteins (I) and				
CC	the polynucleotides (II) that encode them. (I) have cytostatic activity.				
CC	(I) and (II) can be used in gene therapy and vaccine production. (I) and				
CC	(II) may be used in the prevention, diagnosis and treatment of diseases				
CC	associated with inappropriate colon tumour associated protein (TCAP)				
CC	expression, such as colonic cancer. For example, (I) and (II) may be				
CC	used to treat disorders associated with decreased expression by				
CC	rectifying mutations or deletions in a patient's genome that affect the				
CC	activity of TCAPs by expressing inactive proteins or to supplement the				
CC	patients own production of them. Additionally, (II) may be used to				
CC	produce the TCAP proteins, by inserting the nucleic acids into a host				
CC	cell culturing the cell to express the protein. (II) and its				
CC	complementary sequences may also be used as DNA probes in diagnostic				
CC	polymerase chain reaction (PCR) and hybridisation assays to detect and				
CC	quantitate the presence of similar nucleic acids in samples, and				
CC	therefore which patients may be in need of restorative therapy. (I) may				
CC	also be used as antigens in the production of antibodies against TCAPs				
CC	and in assays to identify modulators of TCAP expression and activity.				
CC	Anti-(I) antibodies and antagonists may also be used to down regulate				
CC	TCAP expression and activity. The anti-(I) antibodies may also be used				
CC	as diagnostic agents for detecting the presence of TCAPs in samples				
CC	(e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512				
CC	and AAM24494 to AAM24523 represent nucleotide and amino acid sequences				
CC	given in the exemplification of the present invention.				
XX					
SQ	Sequence 914 AA;				
Query Match		99.9%;	Score 4756;	DB 22;	Length 914;
Best Local Similarity		99.9%;	Pred. No. 0;		
Matches 913;		Conservative	1;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY	1	MGPFKSSVFILHLLEGALSLSLIQNNNGYEGIVADPNVPDETLIQIKDMVTOA	60		
DB		1. mgpfkssvfllhllegalslsliqnnngyegivadpnvpdetliqikdmvta	60		


```
QY 61 SLYLEATCKRKYFKNVAILIPETWTKKADYVRPKLETYKNADVLVAESTPPGNDPEYTE 120
DB 61 slylfeatgkrfyknvailipetwtkkadyvrpkletyknadvlvaestppgndpeyte 120
QY 121 QMNGCEKGERIHLTPDFIAGKLAIEYQPGRAFVHEWAHLRWGVDFEYNNDKFLVLSNG 180
DB 121 qmngcekegerihltpdfdiagkklaiaypggafkafwewahlrgwvdfeyndekfylvlsng 180
QY 181 RIQAVRCSAGITGTNVVKKCGGSCYTKRCTFNKVTGLYEKCEPVLQSRQTEKASIMFA 240
DB 181 riqavrcsagitgtnvvkkcggscytkrctfnkvtgllyekgcefvlsrgtekasimfa 240
QY 241 QHVDSTVECTEQNHKNEAPNOKNCLNRSTWEVTRDSEDFKTKTPMTQPPNPTFSL 300
DB 241 qhvdstivecteqnhkneapnknqkcnlrstwevtrdsedfktktptmtqppnptfsl 300
QY 301 QIGQRIVCLVLDKSGSMATGNRLNRLNQAGQLFLQTVELGWSWGMVTFDSSAAHVQSELI 360
DB 301 qigqrivclvldksgsmatgnrlnrlnqagqlflqtvelgswgmvtfdssahvqselli 360
QY 361 QINSGSDRDTLAKRLPAAASGGTSCSLRSFTVIRKKYPTDGESEIVLLTGDGNTISG 420
DB 361 qinsgsdrdtlakrlpaaasggtscslrsftvirkkypdtgeseivlltgdgntisg 420
QY 421 CFNEVKQSGAIIHTVALGPSAAQELLSKMTGGIQTVASDOVQNGLLIDAFGALSSNG 480
DB 421 cfnevksgaaihtvalgpsaaqelelskmtggigtvasdovqngllidafgalsng 480
QY 481 AVSORSIQLESKGLTLQNSQNGMGIYVDSTVGTGKDTFLITWTQPPQILLWDPGQKOG 540
DB 481 avsqrsiqleskgltlqnsqngmgtvdstvgkdtflitwtppqillwdpsgqkag 540
QY 541 GFVVDKNTKMYLQIPGIAKVTWYSLQASSQTLTLVTSRASNATLPPITVTSKTNKD 600
DB 541 gfvdvdkntkmylqipgiakvtkyslqassqtltlvtsrasnatlppiltvtsktnkd 600
QY 601 TSKFSPVLVYVNIHQGASPIILRASVTALIESVNGKTVTLELLDNGAGADATKDDGVYSR 660
DB 601 tskfspvlvnyvnihqgaspiilrasvtaliesvngkvtvlelldngagadatkdgvysr 660
QY 661 YFTYDTNGRYSVKVRALGGVNAARRRVIPOQSGALYIPGWIENDEIOWNPPRPEINKDD 720
DB 661 yfttydtngrysvkvralggvnaarrvipoqsgalyipgwiendeiownpprpeinkdd 720
QY 721 VOHQVCFRTSSGSGFVASDVPNAPIDLPFGQITDLKAIHGSGLINLWTAPGDDY 780
DB 721 vohqvcftrtssgsgfvasdvpnapidlpfgqitdlkaihgsglinlwtapgddy 780
QY 781 DHGTAKHYIIRISTLDLDRKFNSLQVNTTALIPKEANSSEVFLFKPENITFFNGTDL 840
DB 781 dhgtahkyiiristldldrknfnsdqvnnttalipkeanseeflfpknitffngtdl 840
QY 841 FIAQAVDKVDLKSEISNTARVSLFIPQTPPETSPDTSAPCPNIHINSTIPGIHLK 900
DB 841 fiaqavdkvdlkseisnlarvslfipqtppetpspdtspcpnihinstipgihlk 900
QY 901 INWKWIGELQLSIA 914
DB 901 inwkwigelqlsia 914
```

RESULT 5

AAB73716

ID AAB73716 standard; Protein: 914 AA.

XX

AC AAB73716;

XX

DT 11-SEP-2001 (first entry)

XX

DE Human CLCA1 protein, SEQ ID NO:2.

XX

KW Human CLCA1; goblet cell; mouse Gob-5 orthologue; drug screening; expression inhibition; antisense therapy; gene therapy;

KW

```
KW chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic.
XX Homo sapiens.
FH Key Location/Qualifiers
FT Misc-difference 152
FT /note= "Encoded by AGG in AAH46124"
XX WQ200138530-A1.
XX PD 31 MAY-2001.
XX 22 NOV-2000; 2000WO-JP08232.
XX 24 NOV-1999; 99JP-0333479.
XX 27 APR-2000; 2000JP-0127589.
XX (TAKE ) TAKEDA CHEM IND LTD.
XX PI Nakanishi A, Morita S;
XX WPI; 2001-355935/37.
XX N-PSDB; AAH46102, AAH46124.
XX New antisense nucleotide, useful for treatment and prevention of
XX bronchial asthma and chronic obstructive pulmonary disease
PS Claim 2; Page 76-80; 104pp; Japanese.
XX The invention relates to an antisense nucleotide targetted to the mouse
XX Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
XX the CLCA1 gene (coding sequence shown in AAH46102). The invention also
XX relates to an antibody specific for the Gob-5 protein, medical and
XX diagnostic compositions containing the antisense nucleotide or the
XX antibody, and methods and kits for screening for compounds which inhibit
XX the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.
XX The antisense oligonucleotides and antibody are therefore useful for the
XX treatment and prevention of bronchial asthma and chronic obstructive
XX pulmonary disease. The present sequence represents human CLCA1 protein.
XX Sequence 914 AA:
```

```
Query: Match 99.9%; Score 4756; DB 22; Length 914;
Best local Similarity 99.9%; Pred. No. 0;
Matches 913; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGPFKSSVFILILHLLEGALSNSLIQLNNGYEGIVVAIDPNVPDETLIQIKDMVTOA 60
DB 1 mgpfkssvfillhllegalsnslqnnngyegivvaiddpnvpdedtliqqikdmvtqa 60
QY 61 SLYLEATCKRKYFKNVAILIPETWTKKADYVRPKLETYKNADVLVAESTPPGNDPEYTE 120
DB 61 slylfeatgkrfyknvailipetwtkkadyvrpkletyknadvlvaestppgndpeyte 120
QY 121 QMNGCEKGERIHLTPDFIAGKLAIEYQPGRAFVHEWAHLRWGVDFEYNNDKFLVLSNG 180
DB 121 qmngcekegerihltpdfdiagkklaiaypggafkafwewahlrgwvdfeyndekfylvlsng 180
QY 181 RIQAVRCSAGITGTNVVKKCGGSCYTKRCTFNKVTGLYEKCEPVLQSRQTEKASIMFA 240
DB 181 riqavrcsagitgtnvvkkcggscytkrctfnkvtgllyekgcefvlsrgtekasimfa 240
QY 241 QHVDSTVECTEQNHKNEAPNOKNCLNRSTWEVTRDSEDFKTKTPMTQPPNPTFSL 300
DB 241 qhvdstivecteqnhkneapnknqkcnlrstwevtrdsedfktktptmtqppnptfsl 300
QY 301 QIGQRIVCLVLDKSGSMATGNRLNRLNQAGQLFLQTVELGWSWGMVTFDSSAAHVQSELI 360
DB 301 qigqrivclvldksgsmatgnrlnrlnqagqlflqtvelgswgmvtfdssahvqselli 360
QY 361 QINSGSDRDTLAKRLPAAASGGTSCSLRSFTVIRKKYPTDGESEIVLLTGDGNTISG 420
DB 361 qinsgsdrdtlakrlpaaasggtscslrsftvirkkypdtgeseivlltgdgntisg 420
```

Tue Apr 2 09:40:04 2002

Db 361 qinsgsdrtdtlakrlpaaasgtsicglsraftvirkkpytdgseivlltdgedntisg 420
QY 421 CFNEVKSGALIIHTVALGSPAQAQLEELSKMTGSLQTYASDOVONNGLIDAFGALSSGNG 480
Db 421 cfnevksgaliihtvalgspaqaqleelskmtgslqtyasdvonnglidafgalssng 480
QY 481 AVSORSIOLESKGLTLQNSQWMNGTVIVDSTVGKDTLFLITWTQPPQILLWDPSCQKOG 540
Db 481 avsrslqleskgltlqnsqwmngtvivdstvgkdtlflitwtqppqillwdpdqkqg 540
QY 541 GFVVDKNTKMYLQIPGIAKVGWKYSIQASSQTLTLTTSRASNTLPPITVTSKTKD 600
Db 541 gfvdnkntkmylqipgiakvgtwkyslqassqtltltsrsasnatlppitvtsktnkd 600
QY 601 TSKFSPPLVYANIRQASPLIRASVTALIESVNGKTVTLLELDNGAGADATKDDGYISR 660
Db 601 tsfkpsplvvyanirqgasplirastvltaliesvngkvtlleldngagadatkdgyysr 660
QY 661 YFTTYDTNGRYSVKVRALGGVNAARRVIPPQSGALYIPGWIEDEIOWNPPRPEINKDD 720
Db 661 yfttydtngrys vkvr alggvnaarrvipqsgalyipgwiendeiqwnpprpeinkdd 720
QY 721 VOHQVCFESRTSSGGSFVASDPVNAPIPOLFPPOITDLKAEIHGGSILINLTWTAPGDDY 780
Db 721 vqhkvqcfesrtssggsfvasdvpnapipolpfpqgitalkaelhggsilintwtapgddy 780
QY 781 DHGTAHYIIRISTILDRLKFNESLOVNTALIPKEANSEEVFLFKPENITFENGTDL 840
Db 781 dhgtahyiristsildrlkfneslqvnttalipkeanseevflfkpenitfengtdl 840
QY 841 FTAIQAVDKVLDKSETSNIRVSLPIPPOTPPETSPDTSAPCPNIRHINSTIPGIHLK 900
Db 841 ftaiqavdkvldksetsnirvslpippotppetpspdsapcpnirhinstipgihlk 900
QY 901 IMWKWIGELQSLTA 914
Db 901 imkwigelqslta 914
RESULT 6
ID AAG75614 standard; Protein; 869 AA.
XX AC AAG75614:
XX DT 03-SEP-2001 (first entry)
XX DE Human colon cancer antigen protein SEQ ID NO:6378.
XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX KW colorectal carcinoma.
XX OS Homo sapiens.
XX PN WO200122920-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000WO-US26524.
XX PR 29-SEP-1999; 99US-0157137.
XX PR 03-NOV-1999; 99US-0163280.
XX PR (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI: 2001-235357/24.
XX DR N-PSDB; AAH35019.
XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -

PS Claim 11; Page 7851-7854; 9803pp; English.
XX AAI32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing P.
CC Inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX Sequence 869 AA:
SQ
Query Match 94.1%; Score 4476; DB 22; Length 869;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 858; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 53 IKDMVTOASLYLFEATGCKRFEKNVAILPETWTKADYVRPKLETYKNADVLVAESTPP 112
Db 8 irhexaslylfeatgkrfyknvailpetwtkadyyvrpkletyknadvlvaestpp 67
QY 113 GNDEPYTEQMGNGCKGERIHLPDPFIAGKLAEGYPOGRAFVHEWAHLRMVGFDEYNND 172
Db 68 gndepyteqmgngckgerihltpdfiagklaeyppgrafvhewahlrwgfydeynd 127
QY 173 EKFLVSLNGRIQAVRCSAGITGTNVVKKCGGSCYTKRCTFNKVTGLYKCGCFVLSRQT 232
Db 128 ekfyslmgriqavrcsagitgtnvvkkcgggscytkrctfnkvgllyekgcfvlsqrq 187
QY 233 EKASIMFAQHVDSIVFECTEONHNKEAPNOKNCRSTWEVIRSDSEDFKKTPTMTQTP 292
Db 138 ekasimfaqhvdsivfcteqnhnkeapnqkncrlscwevirsdsedfkktptmtqtp 247
QY 293 PNPTFSLLOIGQRIQVCLVLDKSGSMATGNRLNRLNAGOLFLLQTVELSGWGVMTFDSA 352
Db 248 pnptfslloigqrivclvldksgsmatgnrlnrlndagglflitvelgswgvmtfdsa 307
QY 353 AHVQSELIQINSQSDRDTLAKRLPAAASGGTSCSLRSFTVIRKYPPTDGSSEIVLLTD 412
Db 308 ahvqseliqnsgsdrdtlakrlpaaasgtsicglsraftvirkkpytdgseivlltd 367
QY 413 GEDNTISGCFNEVKSGALIIHTVALGSPAQAQLEELSKMTGSLQTYASDOVONNGLIDAF 472
Db 358 gedntisgcfnevksgaliihtvalgspaqaqleelskmtgslqtyasdvonnglidaf 427
QY 473 GALSNGGAVSQRSIQLESKGLTLQNSQWMNGTVIVDSTVGKDTLFLITWTQPPQILLW 532
Db 428 galsnggavsqrsiqleskgltlqnsqwmngtvivdstvgkdtlflitwtcqpqillw 487
QY 533 DPSGOKGGFVVDKNTKMYLQIPGIAKVGWKYSIQASSQTLTLTTSRASNTLPPIT 592
Db 488 dpsgokggfvvdnkntkmylqipgiakvgtwkyslqassqtltltsrsasnatlppit 547
QY 593 VTSKTNKDTSKFSPPLVYANIRQASPLIRASVTALIESVNGKTVTLLELDNGAGADAT 652
Db 548 vtsktnkdtskfsspplvvyanirqgasplirastvltaliesvngkvtlleldngagadat 607
QY 653 KDGYSRYFTTYDTNGRYSVKVRALGGVNAARRVIPPQSGALYIPGWIEDEIOWNPP 712
Db 608 kdgysryfttydtngrys vkvr alggvnaarrvipqsgalyipgwiendeiqwnpp 667

RESULT 8
AAB73715
ID AAB73715 standard; Protein; 913 AA.

AC AAB73715;
DT 11-SEP-2001 (first entry)
DE Mouse Gob-5 protein, SEQ ID NO:1.

XX Mouse Gob-5; murine; goblet cell; human CICAL orthologue; drug screening;
KW expression inhibition; antisense therapy; gene therapy; bronchial asthma;
KW chronic obstructive pulmonary disease; antiasthmatic.

XX Mus sp.
XX WO200138530-A1.
XX 31-MAY-2001.
XX 22-NOV-2000; 2000WO-JP08232.
XX 24-NOV-1999; 99JP-0333479.
XX 27-APR-2000; 2000JP-0127589.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Nakanishi A, Morita S;

XX WPI; 2001-355935/37.
XX N-PSDB; AAH46101, AAH46120.

XX New antisense nucleotide, useful for treatment and prevention of
PT bronchial asthma and chronic obstructive pulmonary disease

XX Claim 1; Page 72-76; 104pp; Japanese.

XX The invention relates to an antisense nucleotide targetted to the mouse
CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
CC the CICAL gene (coding sequence shown in AAH46102). The invention also
CC relates to an antibody specific for the Gob-5 protein, medical and
CC diagnostic compositions containing the antisense nucleotide or the
CC antibody, and methods and kits for screening for compounds which inhibit
CC the protein. Gob-5 and CICAL are proteins expressed by goblet cells.
CC The antisense oligonucleotides and antibody are therefore useful for the
CC treatment and prevention of bronchial asthma and chronic obstructive
CC pulmonary disease. The present sequence represents mouse Gob-5 protein.

XX Sequence 913 AA;

Query Match 76.8%; Score 3656.5; DB 22; Length 913;
Best Local Similarity 75.8%; Pred. No. 2.4e-297;
Matches 694; Conservative 102; Mismatches 112; Indels 7; Gaps 4;

Qy 1 MGPKSSVFILLLHLCALNSLIQLNNGYGVIVADPNVPDETLIQIKDMVTA 60
Db 1 meslkspsvflilhlhlegvlesliqinnngyegivaiadhvpedealqihkdmvtqa 60
Qy 61 SLYLFEATGKRFYFNVAIIETWTKADYVRPKLETYYKNADVLVAESTPPGNDPEYTE 120
Db 61 spylfeatgkrfyfnvaillpesswkwakpeytrpkletfnadvlvtsttspgndpeyte 120
Qy 121 QMNGCGEKGERIHLTPDFIAGKLAIEYGPQGRAFVHSEMAHLRWGVFDEYNDEKFFYLSNG 180
Db 121 hlgacgekgirihltpldflagkktqygpqdrtrfvhewahrfwgvfdeynndekfyslg 180
Qy 181 RTOAVRCSAGITGTNNVKKCOGGSCYTK-RCTFNKVTGLYKCEGFVLSQRQTEKASIMF 239
Db 181 kpgavrcsaaltcgkqvrcggscitngkcvldrvlglykdcvfvdpdqhnekasimf 240
Qy 240 AQHVDISVEFTEQHNKAPNQKQCNLRSTWEVIRDSDFKKTTPMTTPNPFTSL 299

Db 241 nqinsvvefcteknhnqeqndqncnlnrstwvigesdfkgttmtcappaptfsl 300
Qy 300 LOIQORIVCLVLDKSGSMATGNRLNRLNOAQOLFLLQTVELGWSVGMVTFDSAAHVQSEL 359
Db 301 lqigrlivclvldksgsmldnddrlnrmnqasrlflqtveqgswgmvtfdseayvqsel 360
Qy 360 IQINSGSDRDTLAKRLPAAASGGTSCSGRSRFAFTVIRKKYPTDGSSEIVLLTGDGENTIS 419
Db 361 kqlnsagdrdlilkhlpvaaggtscsgrltaftvikkpytdgseivltdgedntis 420
Qy 420 GCFNEVKOSGAIITVALGPSAAOELEELSKMTGGLQTYASDOVQNNGLIDAFGALSNG 479
Db 421 scfdlvksggalintvalgpaaakeleqskmrggltyssdqvnqnglvdaafalssgn 480
Qy 480 GAVSORISQLESKGLTILQNSQMMNGTVIVDSTVGKDTFLITWTTPQPPILLWDPSCQKQ 539
Db 481 aaiqahsiqlsgrvnlqnqnmngsvivdssvgkdtflitwtthphtflwdpsaveq 540
Qy 540 GGFVVDKNTKWAYLOIPGIAKVGTKYSLOASSOTLTTLTTSRASNTLPPITVTSKTNK 599
Db 541 ngfildtttkvaylqvgptakvgfwkysiqassqtlitvtsraasatlppitvtpvnk 600
Qy 600 DTSKFPSPLVVYANIROGASPIIRASVTALIESVNGKVTTLLELDNGAGADATKDDGYIS 659
Db 601 ntgkfpstvtyasirqgaspillrasvtaliesvngkvtlledngagadatkndgyis 660
Qy 660 RYFTTYDNGRYSVKVRALGCVNAARRVIPQOSGALYIPGWIENDEIQWNPPEINKD 719
Db 661 rftafangrysvkwalgvgvtsdrqaapknramyldgwiedgevrmmpprptesy- 719
Qy 720 DVQHKQVCFRSTSSGGSFVASDVP-NAPIPDLFPFPGQITDLKAEIHGSLNLIWTAPGD 778
Db 720 -vqdkqlcfstssggsfvatnpaaapldlfpccqitdkasiggnlvtwtapgd 778
Qy 779 DYDHGTAHYIIRISILDLRDKFNESIQVNTTALIPKEANSEEVFLFKPENITFNGT 838
Db 779 dydhrasnyfirmstsvldrhntslqvnntgtlplkeasseeifeiggnlfngt 838
Qy 839 DLFTATQAVDKVDLKFSEISNIARVSLFIPPQTPPEPSPDETSAPCPNIHINSTIPGIHI 898
Db 839 difiaiqavdknlskeiseiniarvsvfipaqp---pipedstppcpdisinstipgihv 895
Qy 899 IKIMWKWIGELQLSI 913
Db 896 lkimkwkwlgenqvtl 910
RESULT 9
AAB73715
ID AAB73715 standard; protein; 919 AA.
AC AAB73715;
DT 05-APR-2000 (first entry)
DE Membrane-bound protein PRO1124.
KW Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW pharmaceutical; receptor immunoadhesin; gene mapping.
XX Homo sapiens.
XX WO9963088-A2.
XX 09-DEC-1999.
XX 02-JUN-1999; 99WO-US12252.
XX 02-JUN-1998; 98US-0087607.
XX 02-JUN-1998; 98US-0087609.
XX 02-JUN-1998; 98US-0087759.
XX 03-JUN-1998; 98US-0087827.

PR 04-JUN-1998; 98US-0088021.
 PR 04-JUN-1998; 98US-0088025.
 PR 04-JUN-1998; 98US-0088028.
 PR 04-JUN-1998; 98US-0088029.
 PR 04-JUN-1998; 98US-0088030.
 PR 04-JUN-1998; 98US-0088033.
 PR 04-JUN-1998; 98US-0088326.
 PR 05-JUN-1998; 98US-0088167.
 PR 05-JUN-1998; 98US-0088202.
 PR 05-JUN-1998; 98US-0088212.
 PR 05-JUN-1998; 98US-0088217.
 PR 09-JUN-1998; 98US-0088655.
 PR 10-JUN-1998; 98US-0088722.
 PR 10-JUN-1998; 98US-0088730.
 PR 10-JUN-1998; 98US-0088734.
 PR 10-JUN-1998; 98US-0088738.
 PR 10-JUN-1998; 98US-0088740.
 PR 10-JUN-1998; 98US-0088741.
 PR 10-JUN-1998; 98US-0088742.
 PR 10-JUN-1998; 98US-0088810.
 PR 10-JUN-1998; 98US-0088811.
 PR 10-JUN-1998; 98US-0088824.
 PR 10-JUN-1998; 98US-0088825.
 PR 10-JUN-1998; 98US-0088826.
 PR 11-JUN-1998; 98US-0088858.
 PR 11-JUN-1998; 98US-0088861.
 PR 11-JUN-1998; 98US-0088863.
 PR 11-JUN-1998; 98US-0088876.
 PR 12-JUN-1998; 98US-0089090.
 PR 12-JUN-1998; 98US-0089105.
 PR 16-JUN-1998; 98US-0089440.
 PR 16-JUN-1998; 98US-0089512.
 PR 16-JUN-1998; 98US-0089514.
 PR 17-JUN-1998; 98US-0089532.
 PR 17-JUN-1998; 98US-0089538.
 PR 17-JUN-1998; 98US-0089598.
 PR 17-JUN-1998; 98US-0089599.
 PR 17-JUN-1998; 98US-0089600.
 PR 17-JUN-1998; 98US-0089653.
 PR 18-JUN-1998; 98US-0089801.
 PR 18-JUN-1998; 98US-0089907.
 PR 18-JUN-1998; 98US-0089908.
 PR 19-JUN-1998; 98US-0089947.
 PR 19-JUN-1998; 98US-0089948.
 PR 19-JUN-1998; 98US-0089952.
 PR 22-JUN-1998; 98US-0090246.
 PR 22-JUN-1998; 98US-0090252.
 PR 22-JUN-1998; 98US-0090254.
 PR 23-JUN-1998; 98US-0090349.
 PR 23-JUN-1998; 98US-0090355.
 PR 24-JUN-1998; 98US-0090429.
 PR 24-JUN-1998; 98US-0090431.
 PR 24-JUN-1998; 98US-0090435.
 PR 24-JUN-1998; 98US-0090444.
 PR 24-JUN-1998; 98US-0090445.
 PR 24-JUN-1998; 98US-0090461.
 PR 24-JUN-1998; 98US-0090472.
 PR 24-JUN-1998; 98US-0090535.
 PR 24-JUN-1998; 98US-0090538.
 PR 24-JUN-1998; 98US-0090540.
 PR 25-JUN-1998; 98US-0090557.
 PR 25-JUN-1998; 98US-0090676.
 PR 25-JUN-1998; 98US-0090678.
 PR 25-JUN-1998; 98US-0090688.
 PR 25-JUN-1998; 98US-0090690.
 PR 25-JUN-1998; 98US-0090691.
 PR 25-JUN-1998; 98US-0090694.
 PR 25-JUN-1998; 98US-0090695.
 PR 25-JUN-1998; 98US-0090696.
 PR 26-JUN-1998; 98US-0090862.
 PR 26-JUN-1998; 98US-0090863.
 PR 01-JUL-1998; 98US-0091358.
 PR 01-JUL-1998; 98US-0091360.

PR 01-JUL-1998; 98US-0091544.
 PR 02-JUL-1998; 98US-0091478.
 PR 02-JUL-1998; 98US-0091486.
 PR 02-JUL-1998; 98US-0091519.
 PR 02-JUL-1998; 98US-0091626.
 PR 02-JUL-1998; 98US-0091628.
 PR 02-JUL-1998; 98US-0091633.
 PR 02-JUL-1998; 98US-0091646.
 PR 02-JUL-1998; 98US-0091673.
 PR 07-JUL-1998; 98US-0091978.
 PR 09-JUL-1998; 98US-0091982.
 PR 10-JUL-1998; 98US-0092182.
 PR 10-JUL-1998; 98US-0092472.
 PR 20-JUL-1998; 98US-0093339.
 PR 30-JUL-1998; 98US-0094651.
 PR 04-AUG-1998; 98US-0095282.
 PR 04-AUG-1998; 98US-0095285.
 PR 04-AUG-1998; 98US-0095301.
 PR 04-AUG-1998; 98US-0095302.
 PR 04-AUG-1998; 98US-0095318.
 PR 04-AUG-1998; 98US-0095321.
 PR 04-AUG-1998; 98US-0095325.
 PR 10-AUG-1998; 98US-0095916.
 PR 10-AUG-1998; 98US-0095929.
 PR 10-AUG-1998; 98US-0096012.
 PR 11-AUG-1998; 98US-0096143.
 PR 11-AUG-1998; 98US-0096146.
 PR 12-AUG-1998; 98US-0096329.
 PR 17-AUG-1998; 98US-0096757.
 PR 17-AUG-1998; 98US-0096766.
 PR 17-AUG-1998; 98US-0096768.
 PR 17-AUG-1998; 98US-0096773.
 PR 17-AUG-1998; 98US-0096791.
 PR 17-AUG-1998; 98US-0096867.
 PR 17-AUG-1998; 98US-0096891.
 PR 17-AUG-1998; 98US-0096894.
 PR 17-AUG-1998; 98US-0096895.
 PR 17-AUG-1998; 98US-0096897.
 PR 18-AUG-1998; 98US-0096949.
 PR 18-AUG-1998; 98US-0096950.
 PR 18-AUG-1998; 98US-0096959.
 PR 18-AUG-1998; 98US-0096960.
 PR 18-AUG-1998; 98US-0097022.
 PR 19-AUG-1998; 98US-0097141.
 PR 20-AUG-1998; 98US-0097218.
 PR 24-AUG-1998; 98US-0097661.
 PR 26-AUG-1998; 98US-0097951.
 PR 26-AUG-1998; 98US-0097952.
 PR 26-AUG-1998; 98US-0097954.
 PR 26-AUG-1998; 98US-0097955.
 PR 26-AUG-1998; 98US-0097971.
 PR 26-AUG-1998; 98US-0097974.
 PR 26-AUG-1998; 98US-0097978.
 PR 26-AUG-1998; 98US-0097979.
 PR 26-AUG-1998; 98US-0097986.
 PR 26-AUG-1998; 98US-0098014.
 PR 31-AUG-1998; 98US-0098525.
 PR 16-SEP-1998; 98US-0100634.
 PR 12-JAN-1999; 99US-0115565.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
 PI Wood WI, Yuan J;
 XX
 XX WPI: 2000-072883/06.
 DR N-PSDB; AAZ65095.
 XX
 PT Membrane-bound proteins and related nucleotide sequences -
 PS claim 12; Fig 274; 822pp; English.
 XX
 CC The invention provides membrane-bound PRO polypeptides and

CC polynucleotides encoding them. The PRO sequences of the invention were identified based on extracellular domain homology screening. The PRO sequences have homology with proteins including LDL receptors, TIE ligands and various enzymes. The membrane-bound proteins and receptor molecules are useful as pharmaceutical and diagnostic agents. Receptor immunoadhesins, for instance, can be used as therapeutic agents to block receptor-ligand interactions. The membrane-bound proteins can also be employed for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. The PRO encoding sequences are useful as hybridization probes, in chromosome and gene mapping and in the generation of antisense RNA and DNA. PRO nucleic acid sequences will also be useful for the preparation of PRO polypeptides, especially by recombinant techniques.

XX Sequence 919 AA;

Query Match 60.5%; Score 2879.5; DB 21; Length 919;
Best Local Similarity 61.6%; Pred. No. 3.5e-232;
Matches 563; Conservative 124; Mismatches 216; Indels 11; Gaps 8;

QY 1 MGPPKSSVFILILHLLLEGALNSIQLNNGYEGIVVAIDPNWPEDETLIQIQKDMVTOA 60
DB 1 mgflgrfvllvclllhqs-ntsfiklnngfedividsvpdekileqiedmvtta 59
QY 61 SULFEATGKRPYFNKVAAILPETWTKKADYVRPKLETYKNADVLVAEPSTPPGNDPEYTE 120
DB 60 stylfeatekrffknvslilpenkenpykrpkhenkhadvivapptlgrdepytk 119
QY 121 QMGNGCEGERIHLRPFDFIACKKLAEYQPGQRAVFVHEWAHLRWGVDFEYNDKFXLSNG 180
DB 120 qftecegeqeyihfpidlllgkqneyppgklfhwahlrwgvdfeyndeqpfraks 179
QY 181 -RIQAVRCAGITGVNVYKCKGSGCYTKRGTENKVTGLYKGCCEFLVLSQTEKASIMF 239
DB 180 kkleatrcagisgrnrvykcggscslsracridsttklykdcqffkdkvqtekasimf 239
QY 240 AQHVDIVFECYFQNHKEAPKQKQNLSTWETVRDSEDFKKTPTMTTQPPNPFTSL 299
DB 240 mqsidavvefcnektqngeapsliqnkfrstwevnsedfknktpmvtppppvfsl 299
QY 300 LQGTQIVCLVLDKSSMATNRLNLAQAGOLFQTVELGSGWGVTFDFAAHVQSEL 359
DB 300 lkisqrivclvldksgsmgkdrlnrmnqaakflqtvengswrgvmhfdstativnkl 359
QY 360 IQTNSGSDRTLAKRLPAAASGTSICSLRSFAFTVIRKKY-PTDGEIVLITDGEDNTI 418
DB 360 iqkssderntlmagiplyplgtsicsgikyafigelshsgldgsevlilltdgednta 419
QY 419 SGCFNEVKQSGALIHVVALGPSAAQELSELKMTGGLQTVASDOVONNGLIDAFGALSNG 478
DB 420 sscidevkgsgalvfhialgraadeaviemskitgshrfyvdeaqnnglidaafgaltsg 479
QY 479 NGAVSORSIOLESKGLTLONSOMNGTVIVDSTVVGKDTLFLITWTTPQPOILLWDPGOK 538
DB 480 ntdlsqskqlskleskgltnsnawmdtviidstvgkdtfflitwnslppsisislwdpsgti 539
QY 539 QGGFVVDKNTKMYLIQIPGIAKGVTKWYSLQASS--QTLTLTYSRASNATLPIPIVTSK 596
DB 540 menftvdatskmaylsipgtakvgtwaynlqakanpetlitvtvtsraansvppitvtnak 599
QY 597 TNKDTSKFPSPVLVYANIROCASPIILRASVTALTESVNGKTVTLLELDNGAGADATKDDG 656
DB 600 mnkdvnsfpmpviyaelqgyvpylganvtafiesqnghtevlelldnngagadsfknng 659
QY 657 VYSRYFTYTDNGRYSVKVRALGVNAARRVIPQSGSALYIPGWIENDIOWNPPEI 716
DB 660 vysryftaytengryslkvrhagantarklrpnlraayipgwwvngvgeanpprel 719
QY 717 NKDDVQHKOVCFRSTSGGSFVSDVPNAPIDFLFPPGQITDLKAEIHGSLINLTWTAP 776
DB 720 d-edtqttdledfsrtasgafvsgvpslplpdqyppsqitldatvhedkii-ltwtap 777

QY 777 GDDYDHGTAHYIIRISTSLDLRKNFESLQVNTTALIPKESANSEEVFLFKPENTTFEN 836
DB 778 gdnfdvkvgyiirisasilldrdsddalqvnttdlspkeanskstafkxpeniseen 837
QY 837 GTDLFIQIAQVADKDLKSEISNIRVSLFTPPQTPPE-TSPSPETSAPCPNHIHNSITIPG 895
DB 838 athifiaiksidksnitskvsniaqvtilfpqanpdiidptptptptpdkshns---g 894
QY 896 IHLKIMKMWIGEL 909
DB 895 vniistlvslvsgv 908

RESULT 10

AAB87560

ID AAB87560 standard; Protein: 919 AA.

XX AAB87560;

XX 15-MAY-2001 (first entry)

DE Human PRO1124.

XX Human; PRO protein; mapping.

XX Homo sapiens.

XX W0200116318-A2.

XX 08-MAR-2001.

XX 24-AUG-2000; 2000WO-US23328.

XX 01-SEP-1999; 99WO-US20111.

XX 15-SEP-1999; 99WO-US21090.

XX 07-DEC-1999; 99US-0169495.

XX 09-DEC-1999; 99US-0170262.

XX 11-JAN-2000; 2000US-0175481.

XX 18-FEB-2000; 2000WO-US04341.

XX 18-FEB-2000; 2000WO-US04342.

XX 22-FEB-2000; 2000WO-US04414.

XX 01-MAR-2000; 2000WO-US05601.

XX 03-MAR-2000; 2000US-0187202.

XX 25-APR-2000; 2000US-0199397.

XX 22-MAY-2000; 2000WO-US14042.

XX 05-JUN-2000; 2000US-0209832.

XX (GETH) GENENTECH INC.

XX Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;

XX Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;

XX WPI; 2001-183260/18.

XX N-PSDB; AAF92092.

XX Eighty four nucleic acids encoding PRO polypeptides, useful in

XX molecular biology, including use as hybridization probes, and in

XX chromosome and gene mapping.

XX Claim 12; Fig 70; 278pp; English.

XX The present sequence is a human PRO polypeptide (secreted and transmembrane). The PRO protein, and PRO agonists, PRO antagonists or anti-PRO antibodies are useful for preparation of a medicament useful in the treatment of a condition which is responsive to the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO protein may also be employed as molecular weight markers for protein electrophoresis. The PRO coding sequence has applications in molecular biology, including use as hybridisation probes, and in chromosome and gene mapping.

XX Sequence 919 AA;

Query Match	60.5%;	Score	2879.5;	DB	22;	Length	919;
Best Local Similarity	61.6%;	Pred. No.	3.5e-232;				
Matches	563;	Conservative	124;	Mismatches	216;	Indels	11;
Gaps							
QY	1	MGPFKSSVFILHLLEGALSLSLIQNNNGYEGIVVAIDPNVPEDETLQIQIKDMVTOA	60				
DB	1	mlfgrfvllvcllghs-nlsfklinnngfedlvivdpspedekieqiednvtta	59				
QY	61	SLYLEATCKRYFYKNVAILIPETWTKADYVRPKLETYKKNADVLVAESTPPGNDSPYTFE	120				
DB	60	stylfeatekrffknvslipenwnkpykrphenhkhadvivapptlpgpdepytk	119				
QY	121	QMGNGCEKGERIHLTPDFTAGKLAIEYPOGQRAFWAHKRWGVPEDEYNDEKFFYLSNG	180				
DB	120	qftecegekeythftplllgkknqeyppgkllfvhewahlrkwgvfdeydnedqpfyraks	179				
QY	181	-RIQAVRCISAGITGNVNVKKCGGSCYTKRCYFNKVTGLYKGCERVLQSRQTERASIMF	239				
DB	180	kieatrcsagisgrnrvykcggscslsraclrdsttklygkdcqfkdqvkqtekasimf	239				
QY	240	AQHVDSIVEFCTEQNHKKAPKNQKCNLRSTWIEVIRSEDFKTKPTMTOPPNPFFSL	299				
DB	240	mgsidsavvefcnektngheapslqnkcnfrstweivsnedfknktpmvtppppvpsf1	299				
QY	300	LQIGQRIVCLVLDKSCSMATCNRLNLCAGQLFLLOTVELGSGWGMVTFDSAAHVQSEL	359				
DB	300	lkisgrivclvlksgsmgkdrlnrmnqaakhlqvtvengsgvmvhdstativnkl	359				
QY	360	IQTNGSDRDLAKRLPAAASGGSCTSCSLRSFAFTVIRKYY-PTDGEIVLLTDGEDNTI	418				
DB	360	iqtkssderntlmaglpptylpgtsicsgikyafvigelhsgldgsevlldtgednta	419				
QY	419	SGCFNEVKOSGAIHTVLPGLPSAAQLEELSKMTGGLQTYASDVONNGLIDAFGALSNG	478				
DB	420	sscidevkgsgaivhialgraaadeaviemskitgshfvsdeaqnglidaftaltsg	479				
QY	479	NGAVSORSIQLESKGLTGLNSOMNCTVIVDSTVGKDTLFLITWTTPPOILLWDPGGOK	538				
DB	480	ndlsqsklqleskgltlnsnawmndtviidstvgkdtfltwslppsislwdpstgtl	539				
QY	539	QGGFVVYDKNTKWAYLOIPGIKAVGTWKYSLOASS--QTLTLVTSRASNTLPPITVTSK	596				
DB	540	menftvdatskmaylsipgtakvgtwaynlqakanpetititvtsraansvppitvnaK	599				
QY	597	TKNDSKFSPLVVYANIIRQASPIILRASVTALIESVNGKTVTLELLDNGAGADATKDDG	656				
DB	600	mukdvnsfspmivyaailqgyvvpigantvafiesqnghtevlelldngagadsfkndg	659				
QY	657	VYSRYFTTYDTNGRYSVKVRALGGVNAARRVTPQOSGALYIPGWTENDEIQWNPPEI	716				
DB	660	vysryftaycengrysvkvrhggantarkirpplnraayipgwwvngcieanpprpei	719				
QY	717	NKDDVQHKQVCFSRYSVSGSVASDVPNAPITPDLFPPGQITDLKABIHGSLINLTWAP	776				
DB	720	d-edtqttdlfdrtasgafvsvqpslpldqyppsqitdlldatvheklii-lwtap	777				
QY	777	GDDYDGTAKHYIRISTSLDRKFNESLQWNTTALLPKKEANSDEVLPKPEPTFEN	836				
DB	778	gdnfdvgkvqrylirisailldrdsfdalgvnttdlspkeanskesfafkpeniseen	837				
QY	837	GTDLFIAQAVDKVLKLSISNIARYSLFIPPTPPE-TPSPDTSAPCPNHIHNTIPG	895				
DB	838	athifaiaksidskntskvsniaqvtilfipqanpddidptptptpdkshns---g	894				
QY	896	IHLTKIMRWIGEL	909				
DB	895	vnistlvsvigsv	908				

RESULT 11
AAB65272
ID AAB65272 standard; Protein; 919 AA.
XX

RESULT 11
AAB65272
ID AAB65272 standard; Protein; 919 AA.
XX

AA65272;	
XX	02-APR-2001 (first entry)
XX	Human PRO1124 (UNQ562) protein sequence SEQ ID NO:379.
XX	Human; secreted and transmembrane protein; PRO; cytosstatic;
KW	cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW	diagnostic assay.
XX	Homo sapiens.
XX	WQ200073454-A1.
XX	07-DEC-2000.
XX	30-MAR-2000; 2000WO-US08439.
XX	02-JUN-1999; 99WO-US12252.
XX	23-JUN-1999; 99US-0141037.
XX	07-JUL-1999; 99US-0143048.
XX	20-JUL-1999; 99US-0144758.
XX	26-JUL-1999; 99US-0145698.
XX	28-JUL-1999; 99US-0146222.
XX	17-AUG-1999; 99US-0149396.
XX	15-SEP-1999; 99WO-US21090.
XX	15-SEP-1999; 99WO-US21547.
XX	08-OCT-1999; 99US-0158663.
XX	30-NOV-1999; 99WO-US28313.
XX	01-DEC-1999; 99WO-US28301.
XX	16-DEC-1999; 99WO-US30095.
XX	20-DEC-1999; 99WO-US30911.
XX	05-JAN-2000; 2000WO-US00219.
XX	06-JAN-2000; 2000WO-US00376.
XX	11-FEB-2000; 2000WO-US03565.
XX	18-FEB-2000; 2000WO-US04341.
XX	22-FEB-2000; 2000WO-US04414.
XX	24-FEB-2000; 2000WO-US04914.
XX	24-FEB-2000; 2000WO-US05004.
XX	02-MAR-2000; 2000WO-US05841.
XX	15-MAR-2000; 2000WO-US06884.
XX	20-MAR-2000; 2000WO-US07377.
XX	(GETH) GENENTECH INC.
XX	Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
XX	Frerara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
XX	Grimaldi CJ, Gurney AL, Kljavin IJ, Napier MA, Pan J, Paoni NF;
XX	Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
XX	Zhang Z;
XX	WP; 2001-032160/04.
XX	N-PSDB; AAF44241.
XX	PRO polynucleotides used to produce polypeptides used to target
XX	bioactive molecules such as toxins, radiolabels or antibodies, to
XX	specific cells, to cause targeted cell death -
XX	Claim 12; Fig 274; 935pp; English.
XX	The present invention describes human secreted and transmembrane PRO
XX	proteins. The PRO proteins have cytosstatic activity. The PRO proteins
XX	can be used for targeted delivery of bioactive molecules, such as
XX	toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
XX	sequences, and their fragments, can be used as hybridisation probes, in
XX	chromosomal and gene mapping, and in the generation of anti-sense RNA
XX	and DNA. They may also be used to produce transgenic animals which are
XX	used to develop and screen therapeutically useful reagents. The PRO
XX	nucleotide and protein sequence can be used for tissue typing and in
XX	treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
XX	AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
XX	in the isolation of human PRO sequences. AAF44087 to AAF44269 and
XX	AAF65154 to AAB65300 represent human PRO polynucleotide and protein

sequences given in the exemplification of the present invention.

CC	Query Match	60.5%;	Score 2879.5;	DB 22;	Length 919;	
XX	Best Local Similarity	61.6%;	Pred. No. 3.5e-232;			
SQ	Matches 563;	Conservative	124;	Mismatches 216;	Indels 11;	Gaps 8;
	Sequence	919 AA;				
QY	1	MGPKSSVFLILLHLLCALSNLSLIOLNNGYEGIVVAIDPNVPEDETLQQIKDMVTOA	60			
DB	1	mgllrgfvllivcllqhs-ntsfikinnngfedivivdpsvpedekileqiedmvtta	59			
QY	61	SLYLFEATGRFYPKNVAILIPETWTKYKADYVRPKLETYNADVLYAESTPPGNDSEYTE	120			
DB	60	styleatekrffknvillipenkenpkykrpkhenkhadvivaptilpgrdepytk	119			
QY	121	QMGNGEGERIHLTPDPFAGKLAEGYPOGRAFEVHWAHLRWGVEFYNNDEKFEYLSNG	180			
DB	120	qtecegeyihfupdlilgkknegvppgklfhwahlrwgvfdeydnedqpfyraks	179			
QY	181	-RIOAVRCSAGITGNVVKCOGSCYTKRCTFNKVTGLYKGCFFVLOSROTEKASIMF	239			
DB	180	kkeiatrcsaglsgrnrykcgqscisracridstklkgdcqfpdkvqtekasimf	239			
QY	240	AQHVDSIVEFCTEQHNKEARPNOKNLRSTWEVIRDSDFKTTPTMTPTPPNPTFSL	299			
DB	240	mqdsivsvfecnekchngeapslgnknfrstwevinsedfknktipmtppppvfl	299			
QY	300	LOIGORIVCLVDKSGMATNRLNRLNOAGOLFLLQFVLGWSVGMVTFDSAIAHVOSEL	359			
DB	300	lkisqivclvldksgmggkrlrnmqaakhflqtvgngswgmvhfdstatiwnkl	359			
QY	360	IQINGSDRDTLAKRLPAAASGTSICSLRSFAFTVRKKY-PTDGSFVILLTDGEDNTI	418			
DB	360	iqikssderntlmagiltpylggtscigikyafvgihelshqldgsevlldtgednta	419			
QY	419	SGCFNVEKOGALIHVLPASAAQELFELSKMTGGLTYASDQVQNNGLIDAFGALSSG	478			
DB	420	sscidevkgsgaivfhialgraaadeaviemskitgshfyvsdeaqnnglidafgaltsg	479			
QY	479	NGAVSRSQLESKGLTLQNSQWNGTIVDSTVGKDTLFLITWTPTOPPIILLWPPSGOK	538			
DB	480	ntdlsqskleskgitlunsawmndvldstvgkdtffliltwnslpssislwpsgti	539			
QY	539	QCGFVVDKNTKMAYLQIPGIAKVTGWKYSIQASS--QTLTLVTSRASNTLPPITVTSK	596			
DB	540	menftvdatckmaylsipgtakvgtwaynlqakanpetititvtstraansvppitvna	599			
QY	597	TNKDTSKFPSPVIVYANIRGASPIILRASVTALIESVNGKTVTLELLDNGAGADATKDDG	656			
DB	600	mndvnsfpmiavayaelqgyvpyvganvtafiesqnghtevlelldngagadsfkndg	659			
QY	657	VYSRYFTYDTNGRYSVKVRALGVNAARRVVIPOOSGALYIPGWTEDEIOWNPPRPEI	716			
DB	660	vysryfaytengryslkvrhagantarklppinraayipgwwvngaeieanpprei	719			
QY	717	NKDDVQHKQVCFSTSSGGSFVADVPNAPIDLPFPQGITDLKAEIHGSLNLNTWTAP	776			
DB	720	d-edtqtldedfsrtasggaivsvpslpdpqypsqitldatvhekdlil-lwtap	777			
QY	777	GDDYDHTFAHKYIIRISTSLDLRDKFENESLQVNTALIPKEANSEVEFLRPENTTFEN	836			
DB	778	gdnfdgkvqvyriarasilldrdsfdalqynttdlspkeanskesafkpeniseen	837			
QY	837	GTDLFIAIQAQVDKDLKSEISNIARVSLFIPQTPPE-TFSPDETAPCNHINSTIPG	895			
DB	838	athifialksidknltskvsniaqvtilfipqanpddidtpptptpdkshns---g	894			
QY	896	IHLKIMMKWIGEL	909			
DB	895	vnislctivsvigsv	908			

RESULT 12

ID	AAG73854	standard; Protein; 552 AA.
XX		
AC	AAG73854;	
XX		
DT	03-SEP-2001	(first entry)
XX		
DE	Human colon cancer antigen protein SEQ ID NO:4618.	
XX		
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection; colorectal carcinoma; chromosome 1.	
OS	Homo sapiens.	
PN	WO200122920-A2.	
XX		
PD	05-APR-2001.	
XX		
PF	28-SEP-2000; 2000WO-US26524.	
XX		
PR	29-SEP-1999; 99US-0157137.	
PR	03-NOV-1999; 99US-0163280.	
XX	(HUMA-) HUMAN GENOME SCI INC.	
PI	Ruben SM, Barash SC, Birse CB, Rosen CA;	
XX		
DR	WFI; 2001-235357/24.	
DR	N-PSDB; AAH33285.	
XX		
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers -	
PS	Claim 11; Page 6416-6419; 9803pp; English.	
XX		
CC	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patients own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAB7789 represent sequences used in the exemplification of the present invention.	
CC	N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922.	
XX	Sequence	552 AA;
	Query Match	59.5%; Score 2831; DB 22; Length 552;
	Best Local Similarity	99.8%; Pred. No. 1.8e-228;
	Matches 549;	Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	365	GSDRDTLAKRLPAAASGTSICSLRSFAFTVRKKYPTDGSSEIVLLTDGEDNTISGCFNE	424
DB	3	gsdrdtlakripaaasgtsicslrsfaftvrkkypdgsseivlldgedntisgcfne	62
QY	425	VKQSCAIIHTVALGPSAAQELFELSKMTGGLQTVASDQVQNNGLIDAFGALSSGNAVSV	484
DB	63	vkqsgaiihtvalgpsaaqeleelskmtgglqtvasdqvqnnglidafgalssngnavsq	122
QY	485	RSIQLESKGLTLQNSQWNGTIVDSTVGKDTLFLITWTPTOPPIILLWDPSPQKGGFVV	544


```

Db 123 rsiqlskgltlqnsqmmngtviwvdstvgkdlflitwtctppqllwpsqgkggfvv 182
Qy 545 DKNTMAYLIQPIGIAKVGTKWYSLQASSQTLTLVTSTRASNATLPPITVTSKTKNDTSKF 604
Db 183 dkntkmayliqpiakvgtwkyslcassqtlitvtstrasnatlppitvtsktnktskf 242
Qy 605 PSLPVVYANIROGASPIIRASVTALIESVNGKTVTLELLDNGAGADATKDDGVYSRYFTT 664
Db 243 pslpvvyanirgaspilrasvtaliesvngkvtvtlelldngagadatkdvgvysryftt 302
Qy 665 YDTNGRYSVKVYALGVNAARRVPIQSGALYIPGWIEDEIQWNPPEINCKDDVOHK 724
Db 303 ydtngrysvkvralgvgvnaarrvlpqsgalypgwiendeiwnpprpeinkddvqhk 362
Qy 725 QVCFRTSSGGSFVASDVPNAPIDLPFPQGITDLKAEITHGSLINLTWTAPGDDYDGT 784
Db 363 qvcfertsaggsfvasdvpnapidlpfpqgqitdlkaeihtgslinltwtapgdgydgt 422
Qy 785 AHKYIIRISTSLDLRDKFNESLQWNTTALIPKEANSEEVFLFKPENITFENGDLFIAT 844
Db 423 ahkyiiristsildlrkfneslqwnnttalipkeanseevflfkpetitfengtdlfiat 482
Qy 845 QAVDKVDLKSSEISNIARVSLFIPPTPPETSPDTSAPCPNIHINSTIPGTHILKIMWK 904
Db 483 qavdkvdikseisniarvslfippptppetpspdtapcpnihinstipgthilkimwk 542
Qy 905 WTGELQLSLA 914
Db 543 wigelqlsls 552

```

RESULT 13

AA41036
ID AA41036 standard; protein; 943 AA.

AC AA41036;

DT 07-DEC-1999 (first entry)

DE Human lung tumor antigen L762P.

KW Human; lung tumor; lung cancer; T cell stimulation.

OS Homo sapiens.

PN W09947674-A2.

PD 23-SEP-1999.

XX 17-MAR-1999; 99WO-US05798.

XX 18-MAR-1998; 98US-0040802.

XX 18-MAR-1998; 98US-0040984.

XX 27-JUL-1998; 98US-0123912.

XX 27-JUL-1998; 98US-0123933.

XX (CORI-) CORIXA CORP.

XX Reed SG, Wang T;

XX WPI; 1999-571839/48.

XX N-PSDB; AA224653.

PT New isolated lung tumor polynucleotides, used to develop products for

PT the treatment, prevention and monitoring the progression of lung cancer

XX Example 3; Page 137-139; 148pp; English.

XX The invention provides isolated human lung tumor nucleic acids and

CC polypeptides. The polypeptides can be used for the treatment of lung

CC cancer. The polypeptides and polynucleotides can be used to stimulate T

CC cells or antigen presenting cells for use in the treatment of lung
CC cancer. The polypeptides and monoclonal antibodies specific for the
CC polypeptides can also be used to inhibit the development of lung cancer.
CC Agents which bind the polypeptides can be used for detecting lung cancer
CC and for monitoring the progression of lung cancer.

XX Sequence 943 AA;

Qy 2 GPFKSVFLLIHLLEGALSNSL-----IQLNNNGYEGIVVAIDNPVPEDETLIOQIKD 55
Db 8 gpicalnkfvcll---valsselpflgagvqlqngynglltainpvpengnlisnike 63
Qy 56 MYTOASLYLFEATGTRFYEKFNVAIIPETWTKADYVRPKLETYKNADVLVAESTPPGND 115
Db 64 miteasfylnfackrrvfrfnkllipacwkan-nnsklkqesyekanvltvdwygahgd 122
Qy 116 EPTYEQMGNCBKGRIHLTPDFIAGKKL-AYGPGOGRAFWHEWAHLRWGVDFEYNNDEK 174
Db 123 dpytlqy-qcgkqkyihftpnflndnltagysgrvfwahrlwgvdfeynndkp 182
Qy 175 FYLS-NGRIQAVRCSAGITGTWVVKCQGGSCYTKRCTFNKVTGLYEKGEFVLQSRQTE 233
Db 183 fyingngqikvtrcsdltgfv---cekgpcqpcnciisk---lfkegctfynstqna 236
Qy 234 KASIMFAOHVDSIVFECTEQNHKEAPNQNOKNCLNRSWETVRDSEDEKKTTPM--TTQ 291
Db 237 taslmfmsqisvvefcnaasthmqeapnlqmcslrsawdvtadshhsfpmngtel 296
Qy 292 PNPPTFSLQIGORIVCLVLDKSGSNATNRLNRLNQAGQLFLLQTVELGSMVGMVTFDS 351
Db 297 pppptfslveagdkvclvldvsskmaeadrlllqqaaefylmqiveihftvgiasfds 356
Qy 352 AAHVOSELQINSGRDRTLAKRLPAAASGGR--SICSGLRSAFTVIRK-KYPTDSEIV 408
Db 357 kgeiraqlghinsndrkilvsylpttvsaaktdisicgklkgfvevkeingkaygsvml 416
Qy 409 LLTDGDNITSGCFNEVKOSGALIHVVALGPSAAOLEELSKMTGGLQTVASDVQVNNGL 468
Db 417 lvtsgddklngclptvlgssgclhslalgsaaapneelsrltggklfvpdaisnsm 476
Qy 469 IDAFALSSGNGAVSQRISQLESKGLTLQNSQWMMNGTVIVDSTVGKDTLFLITW--TQPP 527
Db 477 idafarissgtdifqghiqilestgenvkphqlkntvtvtdntvgndtmflvtwqagpp 536
Qy 528 QILLWDPSGQK--QGGFVVDKTKMAYLOIPGIKVGTTWKYSL---OASSQTLTLVTGR 582
Db 537 eailfpdgrkyytnnftnltrtaslwipgtakpghwtlyltnthslqlakvtvtr 596
Qy 583 ASNATLPPITVTSKTNKDTSKFSPPLVVYANIROGASPIIRASVTALIESVNGKTVTLEL 642
Db 597 ansavppatveafverdsihphvpmiyankqgfypilnatvttatvpetgdpvtlrl 656
Qy 643 LDNGAGADATKDDGVYSRYFTTVDYNGRYSVKVRLGGLGVNAARRRVIPOQ----SCALYI 698
Db 657 lddgagadvikndgiysryffsfaangryslkvh-----vohspsistpshsipyghamyv 712
Qy 699 PGWIEDEIQWNPPEINCKDDVQHKQVCFSTRSSGGSFVASDVPNAPIDLPFPQGITD 758
Db 713 pyytangnlqmaprkvsgrneeerkw-gfsrvssggsfsvlgvpagphdpvfpckiid 771
Qy 759 LKAEHGGSLINLTWAPGDDYDGHGTAHYIIRISTSLDLRDKFNESLQWNTTALIPKE 818
Db 772 lea-vkveeeltlswtapedfdqgatsyeirmskslqnldqddfnalvntsknqpqr 830
Qy 819 ANSEEVFLFKPENITFENGTD-----LFTAIQAVDKVDLKSSEISNIARVSLFT 866
Db 831 agirelftfspqlst--nghehqpngetheshryivairamdrnslqgsavsnlaqaplfi 888

Qy	867	PPOTPPETPSD 878	
Db	889	ppnsdp-vpard 899	
RESULT 14			
AA	11321	standard; Protein; 943 AA.	
XX	AA11321		
XX	AA11321		
DT	21-FEB-2001	(first entry)	
XX	Human lung cancer-associated protein L7625.		
XX	Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;		
XX	vaccine; detection.		
OS	Homo sapiens.		
XX	WO200061612-A2.		
PD	19-OCT-2000.		
XX	03-APR-2000; 2000WO-US08896.		
XX	02-APR-1999; 99US-0285479.		
PR	17-DEC-1999; 99US-0466396.		
PR	30-DEC-1999; 99US-0476496.		
PR	10-JAN-2000; 2000US-0480884.		
XX	22-FEB-2000; 2000US-0510376.		
PA	(CORI-) CORIXA CORP.		
XX	PI Wang T, Fan L;		
XX	WPI; 2000-628399/60.		
DR	N-PSDB; AAC65892.		
XX	Isolated polypeptide comprising an immunogenic portion of a lung tumor		
PT	protein is used for detecting and monitoring progression of lung cancer		
PT	in a patient -		
XX	Claim 3; Page 170-172; 261pp; English.		
XX	This invention describes a novel isolated polypeptide (I) which		
CC	comprising an immunogenic portion of a lung tumor protein or variant (P2)		
CC	which have cytostatic activity. The polypeptides and polynucleotides are		
CC	used in compositions and vaccines to inhibit the development of cancer,		
CC	especially lung cancer, in a patient. Methods described in the invention		
CC	can be used to monitor the progression of a cancer by carrying out the		
CC	detection at subsequent time points and comparing the results from the		
CC	different time points. CD4+ and/or CD8+ T-Cells isolated from a patient		
CC	are treated with P2, polynucleotides encoding P2 or antigen presenting		
CC	cells expressing P2 and then administered to the patient to inhibit		
CC	development of cancer.		
XX	Sequence 943 AA;		
XX	Query Match 41.9%; Score 1995.5; DB 21; Length 943;		
XX	Best Local Similarity 45.4%; Pred. No. 5.1e-158;		
XX	Matches 414; Conservative 164; Mismatches 279; Indels 55; Gaps 20;		
Qy	2	GPFFKSVFILLHLLLEGALNSL-----TQLNNGYEGIVVAIDPNVPEFTLQIQKD 55	
Db	8	gpilnlfvtll-----valseilpflgagvqlqdgngylliainpqvpennqlislike 63	
Qy	56	MVTOASLYLEATGKRYFKNVAILPETWTKTRADYVRPKLEYTKNADVLVAESTPPGND 115	
Db	64	mtcasfylvfnatkrvfrfnikilipatwkan-nnskkikqesyekanvittwdyghagd 122	
Qy	116	EPYEQMGCGEKGERIHLTPDFITAGKKL-AEYGPQGRAFVHEWAILRWGVFDEYNDEK 174	

123 dpytlqyrgcgkegyihftpnflnldnltagysgrvfvhewahlrwvfeymndkp 182

175 FYLS-NGRIOAVRCSAGITGTNNVVKCGQSCYTKRCTFNKVTGLYKCEGFEFVLOSQTE 233

183 fyingqngqikvtrcsditgifv---cekpccqencisik---lfkegctflynstqna 236

234 KASIMFAQHVDSDIVFCTEQNHNEKPNKQKCNLSTWEVRDSEDFKKTTPM--TTQ 291

237 tasimfmqslssvvefcnaschnqcapnlqmcslrsawdvtidsadfhshfpmngtel 296

292 PPNPTFSLLOIGORTVCLVLDKSGMATGNLNLNAGOLFLLQTVLGSWGVMTFDS 351

297 ppptfslveagdkvvcclvldvsskmaeadrlqlqaaefylmqiveihtfvgiasfds 356

352 AAHVQSEILQINSQSDRDLAKRLPAAASGCT--SICGLRSATVIRK-KYPTDGSIEV 408

357 kgeiraglhqinsndrklvsvlpttvsaktisicgklkgfevveklngkaygsvmi 416

409 LLTGDGNTISCCFNEVKOSGAIHTVALGPSAAQELSELSKMTGGLOTYASDOVQNNGL 468

417 lvtsgdckllgncldptvissgstihslagsaapnleelriltggikffvdpdisnsm 476

469 IDAFGALSSGNGAVSQRSIQLESKGLTLQNSQMMNGTVIVDSTVGKDTLFLITW--TTOPP 527

477 idafarisgtdifqghiqlestgenkvphqlkatvtdntvgtndtmflvtwgasgpp 536

528 QILLWDPGQK--QGFVVVDKNTKWAYLOIPGIAKVGWTKYSL---QASSOTLTUTVTSR 582

537 eailldpdrkvytnnfitnlfttasiwipgtakpgnhtwtinnthslqalkvtvtsr 596

583 ASNATLPPITVTSKTNKTSKPPSLVVYANIROGASPIRASVTALIESVNGKVTTLLEL 642

597 asnsavppatveafverdsahfphvmyanvkqgfypilnatvtatvepetgdvptlrl 656

643 LDNGAGADATKDDGVYSRYFTTYDNGRYSVKVRALGGVNAARRRVIPOQ-----SGALVI 698

657 lddgagadvikndgiysryffsfaangryslkvh---vnhspstpahslpggshamyv 712

699 PGWIENDELOWNPPEINKDDVOHQVCFRTSSGGSFVSDVPNAPIDPLFPQCQITD 758

713 pgytngnlgmnaprksvgrneerkw-gfsrvsgsgsfvlgvpagbpdvfpckliid 771

759 LKAEIHGGSLLINTWAPGDYDHGTAHYIIRISTSIILDRKFNESLOVNTALIPKE 818

772 lea-vkveeeltlswtapgedfdggdatsyelmkskslnqigddfnallvncskrnpgq 830

819 ANSEEVFLKPENTTPENGTD-----LFATIAQAVDKVDLKSEISNIARVSLFI 866

831 agireiftfppqlst--ngpehpngetheshriyvairamdrnslqgsavsniaqaplf 888

867 PPOTPPETPSD 878

899 ppnsdp-vpard 899

RESULT 15

AA174823

AA174823 standard; Protein; 943 AA.

AA174823

13-JUN-2001 (first entry)

Human ICACC-2 protein sequence.

ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;

interleukin 9 induced calcium activated chloride channel; IL-9;

calcium activated chloride channel; anti-allergic; anti-asthmatic;

anti-inflammatory; immunomodulatory; cystic fibrosis;

inflammatory bowel disease; autoimmune disease.

Homo sapiens.

XX WO9944620-A1.
 XX 10-SEP-1999.
 XX 03-MAR-1999; 99WO-US04703.
 XX 03-MAR-1998; 98US-0076815.
 XX (MAGA-) MAGAININ PHARM INC.
 XX Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;
 XX Nicolaides NC, Zhou Y, Dong Q;
 DR WPI; 1999-550979/46.
 DR N-PSDB; AAF81926.
 XX
 XX New nucleic acid encoding calcium activated chloride channel, used to
 PT identify, e.g. specific modulators for treating atopic allergy -
 XX
 XX ClaIm 13; Fig 4A; 75pp; English.
 XX
 CC The present sequence represents the human interleukin 9 (IL-9) induced
 CC calcium activated chloride channel 2 (ICACC-2) protein. ICACC proteins
 CC have anti-allergic, anti-asthmatic, anti-inflammatory and
 CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
 CC used to alleviate asthma (or more generally atopic allergy), while those
 CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
 CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
 CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
 CC (increased levels) or IBD (reduced levels), also for monitoring
 CC treatment of these conditions. The ICACC proteins can be used:
 CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
 CC reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
 CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
 CC identify modulators and binding partners. ICACC polynucleotides can be
 CC used to generate transgenic animals or recombinant cells, used to screen
 CC for antagonists, also as a source of therapeutic antisense agents or
 CC diagnostic probes (for quantifying mRNA expression, e.g. for
 CC identification of modulators).
 XX
 SQ Sequence 943 AA;

Query Match 41.7%; Score 1986.5; DB 20; Length 943;
 Best Local Similarity 45.1%; Pred. No. 2.9e-157;
 Matches 411; Conservative 167; Mismatches 279; Indels 55; Gaps 20;

QY 2 GPKSSVFILHLLEGALSNSL-----TOLNNGYEGIVVAIDPNVPEDETILQIOIKD 55
 DB 8 gpicnlkfvll-----valsselpfigagvqlqndngynllainpqvpqnqlsnike 63
 QY 56 MVTQASLYLEAFKGRFYFNVAAILPETWKTADYVRPKLEYKNADVLVAESTPPGND 115
 DB 64 miteasfynlnatkrvrrvffnikilpatwkan-nnskkikesyekanvltvdwyrahgd 122
 QY 116 EPTYEQMGNGCEGERHLLTPDFIAGKLL-AEYGPQGRAFVHEWAHLRWGVDFEYNNDEK 174
 DB 123 dpytlgyrcgckegkyihftlfnldnltagysgrvfvhewahlrwgvdfeynndkp 182
 QY 175 FYLS-NGRIQAVRCAGITNTNVKKCGGCVTKRCKTFNKVLEKGEFVLQSRQTE 233
 DB 183 fyingnqqlkvtrcssditgfv---cekgpcqnciisk---lfkcgctfiynstqsa 236
 QY 234 KASIMFAQHVDSEIVFECTEQNHKNEAPNKQKCNLRSTWEVIRDSDFKKTTPM--TTQ 291
 DB 237 tasimfmrslssvvfckasthqeapnlqnqmcslrsawdvtidsadfhhsfpmngtel 296
 QY 292 PPNPTFSLQIGORIVCLVLDKSGMATGNRLNRLNQAGQLFLLQTVELGSLWVGMVTFDS 351
 DB 297 pppptfslveagdkvvclvidasskmaeadrlqlqqaafymqivehtfvqiasfds 356
 QY 352 AAHVQSELQINSGSDRDLTKLPLAAASGGT--SICSLGLRSFAFTVIRK-KYPTDGEIV 408

Search completed: March 30, 2002, 04:17:52
 Job time: 4550 sec

DB 357 kgeiraglhinsndrklvslptvtvsaaktdoisicsglkkkgfvevkeingkaygsvmi 416
 QY 409 LLTDGEDNTISGCFNEVKQSGALIIHTVALGPSAAQEELELSKMTGGLQTYASDQVONNGL 468
 DB 417 lvtsgddkllgnclptvslssgtihsiagsaapnlhelsrlitgglkffvpaidsnsm 476
 QY 469 IDAFGALSSGNGAVSORSTQLESKGLTQLNSQMMNNTVIVDSTVGKDTLFLITW--TTQPP 527
 DB 477 idafsrissgtgdfiqhiglestgenvkphqikntvtvntvgndtmflvtwqasgpp 536
 QY 528 QILLWDPGOK--QGGFVVVDKNTKMAVLOIPGTAKVCTWKYSLOA---SSQTLTLTVTSR 582
 DB 537 eillfpdgrkrytnnftlnlfrtaslwpgtakpghwtynlkhshlqalkvtvtsr 596
 QY 583 ASNATLPPITVTSKTNKTSKFPSPVLVYANIRQASPILRASVTALIESVNGKTVTLEL 642
 DB 597 asnsavppatveafverdslhfpbpmiyanvkqgfypllnatvtatvepetgdpvtlrl 656
 QY 643 LONGAGADATKDDGVYSRYFTTYDTNGRYSVKVRALGGVNAARRRVIPOQ----SGALYI 698
 DB 657 lddgagadvikndgiysryffsaangryslkvh-----vnhspsiatpahslpgshamyv 712
 QY 699 PGWIENDEIQWNPPEINKNDDVOHQVCFSRTSSGGSFVASDVNPAPIPLDFPPGQITD 758
 DB 713 pyytangnlqmaprksvgrne-eernwgfsvrsvsgsfvlgvpagphpdvfpckild 771
 QY 759 LKAEIHGSLINLTWAPGDDYDHGTAHYIIRISTSIILDRDKFNESLQVNTTALIPKE 818
 DB 772 lea-vkveceeltiswcpagedfdggdatsyeirmskslqniqddfnailvntskrnppq 830
 QY 819 ANSEEVFLKPEINITEGTD-----LFTAIQAVDKVLDKSELSINRIARSLFI 866
 DB 831 agirelftspqist--ngpehqpngetheshriyvairamdrnsiqsavnsiaqapllf 888
 QY 867 PPOTPPETPSPD 878
 DB 899 ppsndp-vpard 899

us-09-049-696-41.rag

Tue Apr 2 09:40:04 2002

This Page Blank (uspto)

333 heLeuLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPhe 349
1029 TCCTGCTGCAGACTGTGGAGCAGGAGCTCTGGGTCGGATGTGTACACCTTT 1078
350 AspSerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySe 366
1079 GACAGTGTGCTTATGTATACAAGGCAACTCAACACAGTTAAACAGTGGTGC 1128
366 rAspArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyT 383
1129 TGACAGACGATCTGCTGATCAAGCACTTACCACACAGTATCTGCAGGAGGGA 1178
383 hrSerIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLys 399
1179 CATCTATATGCTCTGGCCCTTCGGACAGCAUATTCAGTGTATAAGAGAAG 1228
400 TyrProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAs 416
1229 TATCCAACTGATGATCTGAAATTTGCTGCTGACCATGGGAGGAGCAA 1278
416 nThrIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleH 433
1279 CACCATTAGCAGCTGCTTTGACCTGGTGAAGCAGACGGCGGCATCATCC 1328
433 IsThrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluLeuSer 449
1329 ATACAGTGGCCCTGGGACCGGCTCGCGTAAAGAGCTTGAGCAGCTGTCC 1378
450 LysMetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnAs 466
1379 AAAATGACAGAGGCGCTCGACATATCTTCGGATCAGGTTTCAGAACAA 1428
466 nGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValS 483
1429 TGGTCTTCTGATGCTTTCGAGCAGCACTCTCCTCAGGAAATCGCGGATCG 1478
483 erGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSer 499
1479 CTCACCACTCCATCCAGCTGGAGAGCAGGGGAGTTAATCTCCAGAATAAC 1528
500 GlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspTh 516
1529 CAATGGATGAATGGCTCAGTATCGTGACAGCTCGGTGGGCAAGGACAC 1578
516 rLeuPheLeuIleThrTrpThrGlnProProGlnIleLeuTrpAla 533
1579 CTTGTTCTTATCACCCTGACACAGCATCTCTCTACAAATATTATCTGGG 1628
533 spProSerGlyGlnLysGlnGlyPheValValAspLysAsnThrLys 549
1629 ATCCCAAGGGAGTGGAAAAATGGTTTATTACTAGACCAACCACTAAG 1678
550 MetAlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLysTy 566
1679 GTGGCTCATCTCCAAAGTCCCAAGCAGCGCTAAGGTGGCTTTTGGAAATA 1728
566 rSerLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSerArg 583
1729 CAGCATTCACGGAGCTCACAGACTCTCACCTTGACTGTCACTCCCTCCGCTG 1778
583 laSerAsnAlaThrLeuProProIleThrValThrSerLysThrAsnLys 599
1779 CAGCAAGTCTACACTGCCTCCTATTACAGTGAACCCGGTAGTGAATAAG 1828
600 AspThrSerLysPheProSerProLeuValValTyrAlaAsnIleArgG 616
1829 AACACAGGGAAATTCGCCAGCCCTGTAAACAGTGTATGCAAGCATTCGCCA 1878
616 nGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerV 633
1879 AGGAGCCCTCGCTATTCTCAGGGCCAGCGTCACAGCCTTCATTGAATCTG 1928
633 alAsnGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAla 649

[illegible]

seq_name: qb_hlc:AK007466

```
seq_documentation_block:
```

seq_documentation_block	LOCUS	AK007466	2915 bp	mRNA	HTC	05-JUL-2001

DEFINITION Mus musculus 10 day old male pancreas cDNA, RIKEN full-length enriched library, clone:1810012P03, full insert sequence.

184 laValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLysCys 200
|||||
579 CAGTGAGGTGTTACGAGCCATACCAATCAAGTTCGTCGGTGC 628
|||||
201 GlnGlyGlySerCysTyrThrLys...ArgCysThrPheAsnLysValTh 216
|||||
629 CAGGAGGCGATTGTATCACTACGGAAGTGTATTCACAGAGTAA 678
|||||
216 rGlyLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrG 233
|||||
679 GGGACTGTATAAGACAAATGTGTATTGTACCAGATCCACCAAAACG 728
|||||
233 LuLysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPhe 249
|||||
729 AGAAGGCTTCATCATGTTTAAACCAAAATATCAATCTGTGTTGAATTC 778
|||||
250 CysThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLy 266
|||||
779 TGTACAGAAAAAATACAAATCAAGAGCCCAAAATGACCAAAACCAACG 828
|||||
266 sCysAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheL 283
|||||
829 ATGCAATCTCCGAAGCAGTGGGAAGTCATCCAGGAATCTCAGGACTTCA 878
|||||
283 ysLysThrThrProMetThrThrGlnProProAsnProThrPheSerLeu 299
|||||
879 AGCAAAACCATCCCATGACAGCCAGCCACCTGCACCCACCTTCTCACTG 928
|||||
300 LeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlyse 316
|||||
929 CTGCAAAATGGACAAAGAAATGTGTGCTTAGTCTTATGATTCGGGAG 978
|||||
316 rMetAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuP 333
|||||
979 CATGCTGAACGATGCTGCTTGACCAAGTGAATCAAGCAAGCGCGCTTT 1028
|||||
333 heLeuLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPhe 349
|||||
1029 TCCGCTGCGACACTGCGACGAGGATCCCTGGGTCGGATGGTGACCTTT 1078
|||||
350 AspSerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlyse 366
|||||
1079 GACAGTGTGCTATGTACAAAGCGAACTCAAAACAGTTAAACAGTGTGC 1128
|||||
366 rAspArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyT 383
|||||
1129 TGACAGAGATCTGCTGTATCAAGCACTTACCACACAGTATCTCGAGGAGGA 1178
|||||
383 hrSerIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLys 399
|||||
1179 CATCTATATGCTCTGGCTTCGGACACACATTTACAGTGATAAAGAAGAAG 1228
|||||
400 TyrProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAs 416
|||||
1229 TATCCAACTGATGATCTGAAATTTGCTGCTGACCGATGGGAGGACAA 1278
|||||
416 nThrIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleH 433
|||||
1279 CACCATTTAGCAGCTGCTTTGACCTGGTGAAGCAGACAGCGGGCGCTCATCC 1328
|||||
433 isThrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluLeuSer 449
|||||
1329 ATACAGTGGCCCTGGGACCGGCTCCCGCTAAAGAGCTTGACAGCTGTGCC 1378
|||||
450 LysMetThrGlyGly...LeuGlnThrTyrAlaSerAspGlnValGlnAsnA 466
|||||
1379 AAATGACAGAGCGCCCTGCAGACATATCTTCGGATCAGGTTCCAGACA 1428
|||||
466 snGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaVal 482
|||||
1429 ATGGTCTGTGTGCTTTCCAGCAGCTCTCTCCTCAGGAATGGCGCATC 1478
|||||
483 SerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSe 499
|||||

1779 GCTCAGCAGCTCCATCCAGCTGGAGACGAGGAGTAAATCTCCAGATAA 1528
|||||
499 rGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspT 516
|||||
1529 CCAATGGATGAATGGCTCAGTCAGTCGGACAGCTCGTGGCAAGGACA 1578
|||||
516 hrLeuPheLeuIleThrTrpThrThrGlnProGlnIleLeuLeuTrp 532
|||||
1579 CTTGTCTTCTATCACCTGGACAAACGCTCTCTACAATATTTATCTGG 1628
|||||
533 AspProSerGlyGlnLysGlnGlyPheValValAspLysAsnThrLy 549
|||||
1629 GATCCACAGGAGTGGAAACAAATGGTTTATAGTACACACACCTAA 1678
|||||
549 sMetAlaTyrLeuGlnIleProGlyIleAlaLysValClyThrTrpLysT 566
|||||
1679 GGTGGCTACTCCCAAGTCCACGAGCCAGCTAAGTTGGCTTTTGGAAAT 1728
|||||
566 yrSerLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSerArg 582
|||||
1729 ACAGCATTCAGCGAGCTCACAGACTCTACCTTCACTGTCACTCCCGT 1778
|||||
583 AlaSerAsnAlaThrLeuProIleThrValThrSerLysThrAsnLy 599
|||||
1779 GCAGCAAGTGTACACTGCCCTCTTATACAGTGACCCCGGTAGTGAATA 1828
|||||
599 sAspThrSerLysPheProSerProLeuValValTyrAlaAsnIleArgG 616
|||||
1829 GAACACAGGGAATTCACAGCCCTGTAAACAGTGTATGCAAGCATTCGCC 1878
|||||
616 lngAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSer 632
|||||
1879 AAGGAGCTCGCTTATCTCAGGCGCCAGCTCACAGCTTGTGATGAATCT 1928
|||||
633 ValAsnGlyLysThrValThrLeuGluLeuAspAsnGlyAlaGlyAl 649
|||||
1929 GTGAATGGAAAAACAGTAACCTGGAATTTACTGGATACGAGCAGGTGC 1978
|||||
649 aAspAlaThrLysAspGlyValTyrSerArgTyrPheThrThrTyra 666
|||||
1979 CGATGCCACCAAGAATGATGGTGTCTACTCAAGGTTTTTTACAGCTTTTG 2028
|||||
666 spThrAsnGlyArgTyrSerValLysValArgAlaLeuGlyGlyValAsn 682
|||||
2029 ATGCAAAATGTTAGATACAGCGTTAAATATATGGGCTCTGGGAGAGTCACT 2078
|||||
683 AlaAlaArgArgArgValIleProGlnGlnSerGlyAlaLeuTyrIlePr 699
|||||
2079 TCACACAGACAGACAGCAGCACCTCCGAGACAGAGCCATGTACATAGA 2128
|||||
699 oGlyTrpIleGluAsnAspGluIleGlnTrpAsnProProArgProGluI 716
|||||
2129 TGGCTGATTCAGGATGGTGAAGTAAGAATGAACCCACACAGCTCCTGAA 2178
|||||
716 leAsnLysAspAspValGlnHisLysGlnValCysPheSerArgThrSer 732
|||||
2179 CTAGTTAT.....GTTCAACAGACAGCAGCTGTGCTTCAGCAGGACATCT 2222
|||||
733 SerGlyGlySerPheValAlaSerAspValPro...AsnAlaProIlePr 748
|||||
2223 TCAGGGGATGCTTTGTGGCCACCAATGTCCCGCAGCAGCTCCCATTC 2272
|||||
746 oAspLeuPheProProGlyGlnIleThrAspLeuLysAlaGluIleHisG 765
|||||
2273 TGACCTCTTCCACCTCTCAAACTCACTGACCTGAAGCCAGCATCCAAAG 2322
|||||
765 lyGlySerLeuIleAsnLeuThrTrpThrAlaProGlyAspTyrAsp 781
|||||
2323 GGCAGAACCTGGTGAATCTGACGTGGAGCGCTCTCTGGGGATGACTACGAC 2372
|||||
782 HisGlyThrAlaHisLysTyrIleIleArgIleSerThrSerIleLeuAs 798
|||||

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICOR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=rc1-CT0249-240 800-211-h10&t3=2000-08-24&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 14
High quality sequence stop: 598.
Location/Qualifiers

FEATURES

source
1..598
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0249"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 150 a 133 c 130 g 185 t
ORIGIN

alignment_scores:
Quality: 979.00 Length: 198
Ratio: 5.021 Gaps: 0
Percent Similarity: 98.485 Percent Identity: 97.980
alignment_block:
US-09-049-696-41 x BE927204/rev ..
Align seg 1/1 to reverse of: BE927204 from: 1 to: 598

660 ArgTyrPheThrThrTyrAspThrAsnGlyArgTyrSerVallysValar 676
|||||
596 AGGTATTTCACACTTATGACAGCAATGGTAGATACAGTGAATGCGG 547
|||||
676 gAlaLeuGlyGlyValAsnAlaAlaArgArgValIleProGlnGlns 693
|||||
546 GGCTCTGGGAGGAGTTAACGACGCCAGCAGAGTGAATACCCAGCAGA 497
|||||
693 erGlyAlaLeuTyrIleProGlyTrpIleGluAsnAspGluIleGlnTrp 709
|||||
496 GTGGAGCACTGTACATACCTGGTGGATTGAGATGATGAATACATGG 447
|||||
710 AsnProProArgProGluIleAsnLysAspValGlnHisLysGlnVa 726
|||||
446 ATCCACCAAGACCTGAATTAATTAAGATGATGTTCAACACAAAGCAAGT 397
|||||
726 lCySPheSerArgThrSerSerGlyGlySerPheValAlaSerAspValp 743
|||||
396 GTGTTTCAGCAGACATCTCGGGAGGCTCATTTGTGGCTTCTGATGTC 347
|||||
743 roAsnAlaProIleProAspLeuPheProGlyGlnIleThrAspLeu 759
|||||
346 CAATGCTCCCATCTACCTGATCTCTCCCACTGGCCAAATCAGCAGCTG 297
|||||
760 LysAlaGluIleHisGlyGlySerLeuIleAsnLeuThrTrpThrAlaPr 776
|||||
296 AAGCGGAAATTCACGGGGAGCTCATTAATCTGACTTGGACAGCTCC 247
|||||
776 oGlyAspAspThrAspHisGlyThrAlaHisLysTyrIleIleArgIles 793
|||||
246 TGGGATGATTATGACCATGGAACAGCTCACAAGTATATCATTCATATA 197
|||||
793 erThrSerIleLeuAspLeuArgAspLysPheAsnGluSerLeuGlnVal 809
|||||
196 GTACAGATATCTTGTATCTCAGACAAAGTTCATGAATCTCTTCAAGTG 147
|||||
810 AsnThrThrAlaLeuIleProLysGluAlaAsnSerGluGluValPheLe 826
|||||

146 AATACTACTCTCTCATCCCAAGAGCAACTCTGAGGAAGTCTTTTG 98
826 uPheLysProGluAsnIleThrPheGluAsnGlyThrAspLeuPheIleA 843
|||||
97 TTAAACCAAGAAACATTAATTTGAAATGACACAGATCTTTTCATTG 48
|||||
843 lAlleGlnAlaValAspLysValAspLeuLysSerGluIleSer 857
|||||
47 CTATTCAGGCTGTGATAAGTC.GATCTGAACACACAGACATATCA 5

seq_name: gb_est2:BG962080

seq_documentation_block:

LOCUS: BG962080 806 bp mRNA EST 12-JUN-2001
DEFINITION: 60282624F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4981853 5',
mRNA sequence.

ACCESSION: BG962080

VERSION: BG962080.1 GI:14349717

KEYWORDS: EST.

SOURCE: house mouse.

ORGANISM:

REFERENCE 1 (bases 1 to 806)
AUTHORS: NIH-MGC http://mgc.nci.nih.gov/
TITLE: National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL: Unpublished (1999)
COMMENT: Contact: Robert Strausberg, Ph.D.
Email: cgabp@r@mail.nih.gov
Tissue procurement: Jeffrey E. Green, M.D.
CDNA Library prepared by: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10983 row: 0 column: 06
High quality sequence stop: 713.

FEATURES

source
1..806
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 248 a 194 c 183 g 181 t
ORIGIN

alignment_scores:
Quality: 951.50 Length: 261
Ratio: 4.101 Gaps: 2
Percent Similarity: 88.889 Percent Identity: 72.031

alignment_block:

US-09-049-696-41 x BG962080

Align seg 1/1 to: BG962080 from: 1 to: 806

1 MetGlyProPheLysSerValPheIleLeuIleLeuHisLeuLeuIgl 17
|||||
17 ATGGAATCTTTGAAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGGA 66
|||||
17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34
|||||
67 AGGAGTCTCTGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTATGAG 116
|||||
34 lYleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
|||||

[illegible]

seq_name: qb_est1:AW361532

seq_documentation_block:	
LOCUS	AW361532 .625 bp mRNA EST
DEFINITION	OV2-CT0261-261099-011-e03 CT0261 Homo sapiens cDNA, mRNA sequence
ACCESSION	AW361532
VERSION	AW361532.1 GI:6866286
KEYWORDS	EST.
SOURCE	human.

Homo sapiens
 Homo sapiens
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 625)
 HCGP <http://www.ludwig.org.br/ORESTES>.
 The FAPESP/LICR Human Cancer Genome Project
 Unpublished (1999)
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics

```

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV2&t2=QV2-CT0261-261039-011-e03&t3=1999-10-26&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 93
High quality sequence stop: 608.
      Location/Qualifiers
        1..625
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone_lib="CT0261"
          /dev_stage="Adult"
          /notes="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 1996
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
172 a 154 c 150 q 149 t
BASE COUNT

```

alignment_scores:			
Quality:	929.50	Length:	200
Ratio:	4.866	Gaps:	1
Percent Similarity:	95.500	Percent Identity:	92.500

alignment block:

US-09-049-696-41 x AW361532

Align seq 1/1 to: AW361532 from: 1 to: 625

[illegible]

|||||
424 AGCCCTCCCAATCTCTCTGGATCCAGTGGACAGCAAGGTGGC 473
542 PheValValAspLysAsnThrLysMetAlaThrLeuGlnIleProGlyI 558
|||||
474 TTTGTAGTGGCAAAACACCAAAATGGCTTACCTCCAAATCCAGGCAT 523
558 eAlaLysValGlyThrTrpLysTyrSerLeuGlnAlaSerSerGlnThrL 575
|||||
524 TGTAGAGTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAACCT 573
575 euThrLeuThrValThrSerArgAlaSerAsnAlaThrLeuProProfil 591
|||||
574 TGACCTGACTGTACGTCCTCGCTGCTCAATGCTACCTGTCTCAATT 623

seq_name: gb_est1:AW753451

seq_documentation_block: 633 bp mRNA EST 28-APR-2000
LOCUS AW753451 QV2-CT0261-261099-011-d05 CT0261 Homo sapiens cDNA, mRNA sequence.

ACCESSION AW753451
VERSION AW753451.1 GI:7668383

KEYWORDS EST.
SOURCE human.

ORGANISM

REFERENCE
AUTHORS
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

MEDLINE

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=QV2&t2=QV2-CT0261-261099-011-d05&t3=1999-10-26&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 66

High quality sequence stop: 90.

FEATURES

source

1..633

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="CT0261"

/dev_stage="Adult"

/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:

SmaI; A mini-library was made by cloning products derived

from ORESTES PCR (U.S. Letters patent application No. 196

,716 - Ludwig Institute for Cancer Research) profiles

into the puc 18 vector. Reverse transcription of tissue

mRNA and cDNA amplification were performed under low

stringency conditions."

136 a 151 c 161 g 185 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 925.00

Ratio: 4.920

Length: 193

Gaps: 0

Percent Similarity: 97.409 Percent Identity: 94.819

alignment_block:

US-09-049-696-41 x AW753451/rev ..

Align seg 1/1 to reverse of: AW753451 from: 1 to: 633

499 SerGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLysAs 515

588 AGCCAGTGGATGAATGCACAGTATCGTGGACAGCAGCGTGGGAAAGAA 539

515 pThrLeuPheLeuIleThrThrThrGlnProGlnIleLeuLeuLeu 532

538 CACTTTGTGCTTATCACCTGGCAATATGCAGCTCCCAAAATCCTTCTCT 489

532 rPaspProSerGlyGlnLysGlnGlyPheValValAspLysAsnThr 548

488 GGGATCCCACTGGACAGCAAGTGGCTTTGTAGTGGACAAAACACC 439

549 LysMetAlaThrLeuGlnIleProGlyIleAlaLysValGlyThrTrpL 565

438 AAAATGGCTTACCTCCAAATCCCAAGGCATTGCTAAGGTGGCAGCTT 389

565 stYrSerLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSer 582

388 ATACAGTCTGCAAGCAAGCTCACAAACCTTGACCTGACTGTCAAGTCC 339

582 rgAlaSerAsnAlaThrLeuProProIleThrValThrSerLysThrAsn 598

338 GTGGCTCCAATGCTACCTCGCTCCCAATTACAGTGACTTCCAAACGAAC 289

599 LysAspThrSerLysPheProSerProLeuValValTyrAlaAsnIleAr 615

288 AAGGACACCAATTCCTCCAGCCCTCTGGCAGCTTTATCAATATTCG 239

615 gGlnGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGlu 632

238 CCAAGGAGCTCCCAATTCCTCAGGGCCAGTGTACAGCCCTGATTGAAT 189

632 erValAsnGlyLysThrValThrLeuGluLeuLeuLeuAspAsnGlyVal 648

188 CAGTGAATGGAAAAACATACCTTGGAACTACCTGGATATGATATCAGGA 139

649 AlaAspAlaThrLysAspAspGlyValTyrSerArgTyrPheThrThr 665

138 GCTGATGCTACTAAGATGATGCGGTGCTAGTCTATGTATATCAGGAAT 89

665 rAspThrAsnGlyArgTyrSerValLysValArgAlaLeuGlyGlyVal 682

83 TCACACCAATGGTAGATACAGTGTAAAGAGCGGCTCTGGGAGGAGTTA 39

682 snAlaAlaArgArgArgValIleProGln 691

38 ACGCAGCCCGACGAGAGCGGATACCCCGAG 10

seq_name: gb_est1:AW361521

seq_documentation_block:

LOCUS AW361521 657 bp mRNA EST 04-FEB-2000

DEFINITION QV2-CT0261-261099-011-f03 CT0261 Homo sapiens cDNA, mRNA sequence.

ACCESSION AW361521

VERSION AW361521.1 GI:6866275

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV2&t2=QV2-CT0261-
261099-011-f03&t3=1999-10-26&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 123
High quality sequence stop: 656.

FEATURES

source

1..657
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0261"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 175 a 160 c 164 g 158 t
ORIGIN

alignment_scores:

Quality: 924.00 Length: 202
Ratio: 4.763 Gaps: 2
Percent Similarity: 96.040 Percent Identity: 93.069

alignment_block:

US-09-049-696-41 x AW361521 ..

Align seg 1/1 to: AW361521 from: 1 to: 657

388 GlyLeuArgSerAlaPheThrValIleArgLysLysTyrProThrAspG 404
|||||
56 GGGATTTCGATCGTCGATTAATGTGATTAGGAACAATATCTAATCATG 105
404 ly.SerGluIleValLeuLeuThrAspGlyGluAspAsnThrIleSerG 420
|||||
106 GATTCTGTAATTGTGCTGCTGACGGATGGGGAAGACAACTATTAAGTGG 155
420 yCysPheAsnGluValLysGlnSerGlyAlaIleIleHisThrValAla 437
|||||
156 GTGCTTTAACGAGGTCAACAAAGTGTGCTCATCCACACAGTCGCTT 205
437 euGlyProSerAlaAlaGlnGluLeuGluLeuSerLysMetThrGly 453
|||||
206 TGGGGCCCTCGCAGCTCAAGAACTAGAGGAGCTGTCCAAAATCACAGGA 255
454 GlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnAsnGlyLeuIleAs 470
|||||
256 GGTTTACAGACATATGCTCGGATCAAGTTCAAGAACAAATGGCCCTCAT 305
470 pAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValSerGlnArgSerI 487
|||||
306 TGCCTTTGGGGCCCTTTCATCAGGAATGGAGCTGTCTCTCAGCGCTCCA 355
487 leGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSerGlnTrpMetAsn 503
|||||
356 TCCAGCTTGAGAGTAAGGATTAAACCTCCAGAACACCCAGCTGATGAAT 405
504 GlyThrValIleValAspSerThrValGlyLysAspThrLeuPheLeuII 520
|||||
406 GGCACAGTGTGCTGGACAGCACCGCTGGGAAGGACACTTTGTTCTTAT 455
520 eThrTrpThrThrGlnProProGlnIleLeuLeuTrpAspProSerGlyG 537

|||||
456 CACCTGGACAAATGAGCTCCCAAAATCCCTCTCTGGGATCCCAAGTGAC 505
537 lnLysGlnGlyGlyPheValValAspLysAsnThrLysMetAlaTyrLeu 553
|||||
506 AGAAGCAAGGTGGCTTGTAGTGGACAAAACACCAAAATGGCTACCTC 555
554 GlnIleProGlyIleAlaLysValGlyThrTrpLysTyrSerLeuGlnAl 570
556 CAAATCCAGGCATTCGTAAGTTGGCACTTGGAAATACAGTCTGCAAGC 605
570 aSerSerGlnThrLeuThrValThrSerArgAlaSerAsnAlaI 587
606 AAGCTCAAAACCTTTGACCTGACTGTCACGTCCCGTCAATGCTA 655
587 hr 587
656 CC 657

seq_name: gb_est2:BF083270

seq_documentation_block:

LOCUS BF083270 528 bp mRNA EST 18-OCT-2000
DEFINITION RCL-CT0249-110900-214-b01 CT0249 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF083270
VERSION BF083270.1 GI:10877100
KEYWORDS EST.
SOURCE human:
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 528)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6L2=RC1-CT0249-110
900-214-b01&t3=2000-09-11&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 33
High quality sequence stop: 526.
Location/Qualifiers
1..528
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0249"
/dev_stage="Adult"

/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

FEATURES

source

158 a 113 c 127 g 130 t
BASE COUNT
ORIGIN

Tue Apr 2 09:40:07 2002

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1-QV2&t2-QV2-CT0261-
261099-011-e07&t3=1999-10-26&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 88
High quality sequence stop: 613.
Location/Qualifiers
1..621
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0261"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

FEATURES

source

BASE COUNT 166 a 151 c 150 g 154 t

ORIGIN

alignment_scores:

Quality: 907.00 Length: 198
Ratio: 4.774 Gaps: 1
Percent Similarity: 95.960 Percent Identity: 94.444

alignment_block:

US-09-049-696-41 x AW361520 ..

Align .seg 1/1 to: AW361520 from: 1 to: 621

391 SerAlaPheThrValIleAargLysLysTyrProThrAspGlySerGluI 407
19 TCTGCATTACTGTGATTATGAACGATTATTA.ACTGATGATCTGTAAT 67
407 eValLeuLeuThrAspGlyGluAspAsnThrIleSerGlyCysPheAsnG 424
68 TGTGCTGCTGACGGATGGGGAAGACACACTATAAGTGGTGCITTAACG 117
424 luValLysGlnSerGlyAlaIleIleHisThrValAlaLeuGlyProSer 440
118 AGGTCAACATAGTGTGCTCATCCACACAGTCTTGGGGCCCTCT 167
441 AlaAlaGlnGluLeuGlu GluLeuSerLysMetThrGlyGlyLeuGlnT 457
169 GCAGCTCAAGAACTAGAGTGAGCTGTCCAAATGACAGAGGTTTACAGA 217
457 hrTyrAlaSerAspGlnValGlnAsnAsnGlyLeuIleAspAlaPheGly 473
218 CATATGCTTCAGATCAAGTTCAGAACATGGCTCATTTGATGCTTTGGG 267
474 AlaLeuSerSerGlyAsnGlyAlaValSerGlnArgSerIleGlnLeuG 490
269 GCCCTTTTCATCAGGAAATGGAGCTGTCTCTCAGCGCTCCATCCAGCTGA 317
490 uSerLysGlyLeuThrLeuGlnAsnSerGlnTrpMetAsnGlyThrValI 507
318 GAGTAAGGATTAACCTCCAGAACAGCCAGTGGATGAATGGCACAGTGA 367
507 leValAspSerThrValGlyLysAspThrLeuPheLeuIleThrTrpThr 523
368 TCGTGGACAGACCGTGGGAAAGACACTTTGTTTCTTATCACCTGGACA 417
524 ThrGlnProGlnIleLeuLeuLeuTtpAspProSerGlyGlnLysGlnG 540

alignment_scores:

Quality: 911.00 Length: 172
Ratio: 5.297 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.419

alignment_block:

US-09-049-696-41 x BF083270 ..

Align seg 1/1 to: BF083270 from: 1 to: 528

643 LeuAspAsnGlyAlaGlyAlaAspAlaThrLysAspAspGlyValTyrSe 659
12 CTGTAATAGGACAGCTGCTGATGCTACTAAGGATGACGGTGTCTACTC 61
659 rArgTyrPheThrThrTyrAspThrAsnGlyArgTyrSerValLysVala 676
62 AAGGTATTTCACAACTATGACACGGAATGCTAGATACAGTGTAAAGTGC 111
676 rGalaLeuGlyGlyValAsnAlaAlaArgArgValIleProGlnGln 692
112 GGGCTCTGGGAGGAGTTAAACGACGACGACGAGAGTATACCCAGCAG 161
693 SerGlyAlaLeuTyrIleProGlyTyrIleGluAsnAspGluIleGlnTr 709
162 AGTGGACCACTGTACATACCTGCTGCTGATGATGATGATGATGATGAT 211
709 pAsnProArgProGluIleAsnLysAspAspValGlnHisLysGlnV 726
212 CAATCCACCAAGACCTGAAATTAATAGGATGATGTTCAACACAGCAAG 261
726 alCysPheSerArgThrSerSerGlyGlySerPheValAlaSerAspVal 742
262 TGTGTTTCAGCAGAACATCTCCGGAGGCTCATAGTGGCTTCTGATGTC 311
743 ProAsnAlaProIleProAspLeuPheProGlyGlnIleThrAspLe 759
312 CCAATGCTCCATACCTGATCTCTCCACCTGGCCAAATCACCGACCT 361
759 uLysAlaGluIleHisGlyGlySerLeuIleAsnLeuThrTrpThrAlap 776
362 GAAGCGGAAATTCACGGGGCAGCTCTCAATTAATCTCACTTGGACGCTC 411
776 toGlyAspAspTyrAspHisGlyThrAlaHisLysTyrIleIleArgIle 792
412 CTGGGGATGATTATGACCATGCAACAGCTCAACAAGTATATCATTCGA 461
793 SerThrSerIleLeuAspLeuArgAspLysPheAsnGluSerLeuGlnVa 809
462 AGTCAAGATATTCTTGATCTCAGACAGCAAGTTCAATGAATCTCTCAAGT 511
809 lAsnThrThrAlaLeu 814
512 GAATACTACTGCTCTC 527

seq_name: gb_est1:AW361520

seq_documentation_block:

LOCUS AW361520 621 bp mRNA EST 04-FEB-2000
DEFINITION OV2-CT0261-261099-011-e07 CT0261 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW361520
VERSION AW361520.1 GI:6866274

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 621)

HCGP http://www.ludwig.org.br/ORESTES.

TITLE The FAPESP/LICR Human Cancer Genome Project

JOURNAL Unpublished (1999)

CONTACT Simpson A.J.G.

LABORATORY Laboratory of Cancer Genetics

|||||
418 ATGCACCCCTCCCAATCTCTCTGGATCCCGAGTGGACAGAAAGG 467
540 yGlyPheValValAspLysAsnThrLysMetAlaTyrLeuGlnIleProG 557
468 TGGCTTTGTAGTGACAAAAACACCAAAATGGCTACCTCCAAATCCAG 517
557 lytleAlaLysValGlyThrTrpLysTyrSerLeuGlnAlaSerSerGln 573
518 GCATTCTTAAGGTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAA 567
574 ThrLeuThrThrValThrSerArgAlaSerAsnAlaThr 587
568 ACCTTGACTGACTGCTGACGCTCCCGTGGCTTCAATGCTACC 609

seq_name: gb_est2:BF578246

seq_documentation_block: 767 bp mRNA EST 12-DEC-2000
LOCUS BF578246 602093073F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4207386 5',
DEFINITION mRNA sequence.

ACCESSION BF578246
VERSION BF578246.1 GI:11651958
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM9770 row: a column: 19
High quality sequence stop: 679.

FEATURES
Location/Qualifiers
1..767
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4207386"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 235 a 176 c 180 g 176 t
ORIGIN

alignment_scores:
Quality: 873.50 Length: 234
Ratio: 4.179 Gaps: 1
Percent Similarity: 89.316 Percent Identity: 73.077

alignment_block:
US-09-049-696-41 x BF578246 ..

Align seg 1/1 to: BF578246 from: 1 to: 767

1 MetGlyProPheLysSerValPheIleLeuLeuHisLeuGln 17
|||||
23 ATGGAATCTTTGAAGAGCTCTGCTTCTTCTGATCCTCCACCTTCTGGA 72

17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34
|||||
73 AGAGATTCTGAGTGAGTCCCTCATCAACTGAACAACACGCGTATGAGG 122
34 lytleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
|||||
123 GCATGCTCATCGCATAGACACGAGCTGCCGGAAGATGAAGCCCTCAT 172
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
|||||
173 CAACACATAAAGGACATGGTGACTAGGCGCTCTCCATACCTGTTTGAAGC 222
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluT 84
|||||
223 TACAGGAAAAAAGATTTTACTTCAAAAATGTTGCCAT.TTGATTCCCGAGA 271
84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
|||||
272 GCTGGAAGCAAAAGCTGATATATACGAGGCGCAAACTTGAACCTTCAAA 321
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
|||||
322 AACGCTGATGCTCTTGTATCAACAACAGCCCTCTAGGCAATGATGAGCC 371
117 oTyrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134
|||||
372 CTACACCGAACATATAGGAGCATGTGGAGAAAGGGGATCAGGATTCCACC 421
134 eutThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
|||||
422 TGACTCTGACTTCTTCCATGAGTGGGCTCACTCCGATGGGAGTGTATA 471
151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
|||||
472 GACAGGACCTTCTCCATGAGTGGGCTCACTCCGATGGGAGTGTATA 521
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGlnA 184
|||||
522 TGAATCAACAACGAGAGAGTCTTACTTTATCCAAAGGAAACCCCAAG 571
184 laValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLysCys 200
|||||
572 CAGTGAGGTGTTCCAGCAGCCATTACCGGTACAAATCAAGTTCGTCGGTGC 621
201 GlnGlyCysSerCysTyrThrLys...ArgCysThrPheAsnLysValTh 216
|||||
622 CAGGAGGCGCAGTGTATCACCTAACGGAAGTGTCTAATCGACAGAGTAAC 671
216 xGlyLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrG 233
|||||
672 GGGAGCTGTATACAGACCATGTGTTGTACCAGATCCAAACAAAGAG 721

seq_name: gb_est2:BF581041

seq_documentation_block:
LOCUS BF581041 661 bp mRNA EST 12-DEC-2000
DEFINITION 602100173F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4219698 5',
DEFINITION mRNA sequence.

ACCESSION BF581041
VERSION BF581041
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

Percent Similarity: 93.035 Percent Identity: 89.552

alignment_block:

US-09-049-696-41 x AW753449

Align seg 1/1 to: AW753449 from: 1 to: 652

```

384 SerIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLys.T 400
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
53  TCCACGTCGCGATGATATGATCGCATTTATTGTGATTAGTAACCAATT 102
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
400 yrProThr....AspGlySerGluIleValLeuLeuThrAspGlyGluAsp 415
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
103 ATCCCTACTGATGATCTCTGTAATTGTCTGCTGACGGATGGGAAGAC 152
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
416 AsnThrIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleI 432
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
153 AACACTATAAGTGGTGCTTTAACGAGGTCAACATAGTGTGCCCTCAT 202
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
432 eHisThrValAlaLeuGlyProSerAla..AlaGlnGluLeuGluLeu 448
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
203 CCACACAGTCGCTTTGGGGCCCTCTCCAGGCTCAAGAACTAGAGGAGCTG 252
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
449 SerLysMetThrGlyGlyLeuGlnThrTyraLaserAspGlnValcLoAs 465
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
253 TCCAAATGACAGGAGGTTTACAGACATATGCTTCAGATCAAGATTCAAGAA 302
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
465 nAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlav 482
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
303 CAATGGCCTCATGTATGCTTTTGGGGCCCTTTCATCAGGAAATGGAGCTG 352
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
482 alSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsn 498
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
353 TCTCTCAGCGCTCCATCCAGCTGTAGAGTAAGGATTAACTCCAGAAC 402
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
499 SerGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLysAs 515
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 ACCCAGTGGATGAATGGACACATGATCGTGGACAGCACCGTGGGAAGA 452
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
515 pThrLeuPheLeuIleThrTrpThrThrGlnProGlnIleLeuLeuLeu 532
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
453 CACTTTGTTCTTATACCTCGACAAATGACAGCTCCCAAAATCCTTCTCT 502
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
532 rpAspProSerGlyGlnLysGlnGlyGlyPheValValAspLysAsnThr 548
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
503 GGGATCCCACTGGACAGAAAGAGGTTGGCTTTGTAGTGACAAAACACC 552
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
549 LysMetAlaTyrrLeuGlnIleProGlyIleAlaLysValGlyThrTriply 565
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
553 AAAATGGCTAGCTCCAAATCCAGGCATTGCTAAGTTGGCACTTGGAA 602
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
565 sTyrrSerLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSer 581
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
603 ATACAGTGTCAAGCAAGCTCACTAACTTGACCTTGACCTGCTACGCTCC 651

```

seq_name: gb_est2:BG963539

seq_documentation_block:

```

LOCUS      BG963539          745 bp      mRNA      EST      12-JUN-2001
DEFINITION 602831217F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4985741 5',
            mRNA sequence.
ACCESSION  BG963539
VERSION    BG963539.1 GI:14351176
KEYWORDS   EST.
SOURCE     house mouse.
            Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
            1 (bases 1 to 745)
AUTHORS    NIH-MGC http://mgi.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.

```

Email: coapbs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LIAM10994 row: a column: 06
High quality sequence stop: 745.
FEATURES
 source
 1..745
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4985741"
 /clone_lib="NCI_CGAP_Co24"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; cloned unidirectionally. Primer: Oligo dT. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 236 a 173 c 172 g 164 t
ORIGIN

alignment_scores:

 Quality: 856.50 Length: 246
 Ratio: 3.807 Gaps: 3
Percent Similarity: 91.463 Percent Identity: 71.951

alignment_block:

US-09-049-696-41 x BG963539

Align seg 1/1 to: BG963539 from: 1 to: 745

```

1 MetGlyProPheLysSerValPheIleLeuIleLeuHisteuLeuGl 17
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
15 ATGGGAATCTTTGAAGAGCTGCTGCTTCTTGTATCTCCACCTCTCTGA 64
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrrGluG 34
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 AGGAGTTCTGAGTGAGTCCCTCATCTCACTGACACACACGGCTATGAGG 114
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
34 lylleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
115 GCATCTGTCATGCCATAGACACGAGCTGCCGGAAGATGAAGCCCTCAT 164
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrrleuPheGluA 67
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
165 CAACACATAAAGACATGGTGACTCAGGCTCTCCATACCTGTTTGAAGC 214
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
67 aThrGlyLysArgPheTyrrPheLysAsnValAlaIleLeuIleProGlu 84
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
215 TACAGGAAAAGATTTTACTTCAAAAATGTTGCCATTTTGTATCCCGAGA 264
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
84 hrTrpLysThrLysAlaAspTyrrValArgProLysLeuGluThrTyrrLys 100
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
265 GCTGGAAGCAAGACGCTGAATATACGAGGCCAAACCTTGAACCTTCAAA 314
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
315 AACGCTGATGTCTCTGTATCAACACACCGCCCTCTAGGCAATGATGAGCC 364
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 oTyrrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHis 133
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
365 CTACACCGACATATAGAGGATGTGGAGACAAAGGGGATCAGGATTCAC 414
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134 LeuThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrrGlyProGl 150
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
415 CTGACTCTCTGACTCTCTAGCAGGAAAGAGCTGACTCAGTATGGGCCACA 464
   ||| |||::: ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
150 nGlyArgAla..PheValHisGluTrpAlaHisLeuArgTrpGlyValPhe 166

```

Tue Apr 2 09:40:07 2002

```

165 AGACAGGACCCCTTTGTCATGAGTGGGCTCACATCCGATGGGAGTGT 514
167 AspGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleG1 183
515 AATGAATACAAAAACGACGAGAGTTCTACTTATCCAAAGGAAAACC.CA 563
183 nAlaValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLysC 200
564 AGCAGTGGGTGTACAGCAGCCATTACCGGTAAACAATCAAGTTCGTGGT 613
200 ysGlnGlyGlySerCysTyrThrLys...ArgCysThrPheAsnLysVal 215
614 GCCAGGGAGGCAG.TGTATCACTAACGGAAAGTGTGTAATCGACAGAGTA 662
216 ThrGlyLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnTh 232
663 ACGGAGCTGTATAA.AGACAATGTGTATTAGTACCAGATCCACACCAAAA 711
232 rGluLysAlaSerIleMetPheAlaGlnHisVal 243
712 CGAAAAAGGCTTCATCATGTGTTAAACCAAAATATC 745

```

OM of: US-09-049-696-41 to: Issued_Patents_NA.* out_format : pfs

Date: Mar 30, 2002 4:12 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+pin_model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09049696/runat_28032002_145435_4695/app_query.fasta_1.987
-BA=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-YGAPOP=6.000 -YGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1_100 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-049-696-41
Query length: 914
Database: Issued_Patents_NA.*
Database sequences: 351203
Database length: 113238999
Search time (sec): 97.350000
```

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-469-667-8	1347.00	2575.79	8.0e-136	878	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCT-US95-07289-8	1347.00	2575.79	8.0e-136	878	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-221-298-34	618.00	1173.62	1.0e-57	401	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-385-982-23	519.00	976.78	9.3e-47	576	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-385-982-25	468.50	878.46	2.8e-41	595	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-385-982-24	447.00	836.27	6.2e-39	618	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-385-982-27	421.50	787.00	3.5e-36	611	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-385-982-33	373.50	691.31	7.4e-31	742	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-617-697-8	139.50	212.19	0.0004	4803	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-409-995-3	137.00	204.91	0.0009	5738	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-685-467-3	137.00	204.91	0.0009	5738	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-913-942-3	137.00	201.62	0.0014	7291	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-038-682-8	131.50	196.97	0.0025	4702	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-302-832-1	131.50	196.97	0.0025	4702	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-530-198-8	131.50	196.97	0.0025	4702	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-880-8	131.50	196.97	0.0025	4702	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-728-470-1	131.50	196.97	0.0025	4702	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-719-641-1	131.50	196.97	0.0025	4702	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-038-682-1	131.00	194.85	0.0033	5116	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-302-832-1	131.00	194.85	0.0033	5116	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-530-198-1	131.00	194.85	0.0033	5116	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-880-1	131.00	194.85	0.0033	5116	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-728-470-1	131.00	194.85	0.0033	5116	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-617-697-1	131.00	194.85	0.0033	5116	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-719-641-1	131.00	194.85	0.0033	5116	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-038-682-5	122.00	169.41	0.0871	9171	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-302-832-5	122.00	169.41	0.0871	9171	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-530-198-5	122.00	169.41	0.0871	9171	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-880-5	122.00	169.41	0.0871	9171	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-728-470-5	122.00	169.41	0.0871	9171	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-617-697-5	122.00	169.41	0.0871	9171	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-719-641-5	122.00	169.41	0.0871	9171	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-103-840A-2	121.50	83.84	5.1e+03	4403765	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-103-840A-1	121.50	83.82	5.1e+03	4411529	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-045-632-1	119.50	174.24	0.0469	4527	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-379-523-4	118.50	176.40	0.0356	3356	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-447-031A-8	118.00	171.07	0.0703	4612	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-891-640-1	118.00	167.12	0.1168	6156	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-592-874-1	117.50	145.01	1.99	28804	1

```
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-096-942-2 117.50 145.01 1.99 28804
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-096-867-2 117.50 145.01 1.99 28804
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-041-886-24 117.00 169.15 0.0900 4608
/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCT-US94-05277-1 117.00 169.15 0.0900 46
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-038-682-6 117.00 159.49 0.3106 9323
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-469-667-8

```
seq_documentation_block:
Sequence 8, Application US/08469667
Patent No. 5733748
GENERAL INFORMATION:
APPLICANT: Yu, Guo-Liang
APPLICATOR: Rosen, Craig
TITLE OF INVENTION: Colon Specific Genes and Proteins
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
ADDRESSEE: Stewart & Olstein
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: NJ
COUNTRY: USA
ZIP: 07068-1739
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,667
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Ferraro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-435
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 878 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 2..685
US-08-469-667-8
```

alignment_scores:

Quality: 1347.00 Length: 258
Ratio: 5.241 Gaps: 1
Percent Similarity: 99.612 Percent Identity: 99.225

alignment_block:

US-09-049-696-41 x US-08-469-667-8

Align seg 1/1 to: US-08-469-667-8 from: 1 to: 878

```
657 ValTyrSerArgTyrPheThrThrTyrAspThrAsnGlyArgTyrSerVa 673
|||||
2 GTCTACTCAAGGTATTTCACAACTTATGACACGACGATGCTAGTACAGTGT 51
673 LlysValArgAlaLeuGlyGlyValAsnAlaAlaArgArgValIleP 690
|||||
52 AAAAGTGGGGCTCTGGGAGGAGTAAACGACGACGACGAGAGTGATAC 101
690 roGlnGlnSerGlyAlaLeuTyrIleProGlyTyrIleGluAsnAspGlu 706
|||||
```

Tue Apr 2 09:40:06 2002

```
102 CCAGCAGAGTGGAGCACTGTATACCTCGCTGGATTGAGATGATGAA 151
707 IleGlnTrpAsnProProArgProGluLeuLeuAsnLysAspValGlnHi 723
152 ATACATGGAATCCACAGACCTGAAATTAATAAGGATGATGTTCAACA 201
723 sLysGlnValCysPheSerArgThrSerSerGlyGlySerPheValAla 740
202 CAAGCAAGTGTGTTTCAGCAGAACATCTCCGGAGGCTCATTTGGGCTT 251
740 erAspValProAsnAlaProIleProAspLeuPheProGlyGlnIle 756
252 CTGATGTCCTCCAAATGCTCCCATACCTGATCTCTCCACCTGGCCAAATC 301
756 ThrAspLeuLysAlaGluLeuHisGlyGlySerLeuIleAsnLeuThrTr 773
302 ACCGACCTGAAGCGGAAATTCACGGGGCAGTCTCTATTANTGACTTG 351
773 pThrAlaProGlyAspAspThrAspHisGlyThrAlaHisLysTyrIleI 790
352 GACAGCTCTCTGGGATGATTTATGACCATGGAACAGCTCACAAGTATATCA 401
790 leArgIleSerThrSerIleLeuAspLeuArgAspLysPheAsnGluSer 806
402 TTCGAATAAGTACAAATGATTTCTGATCTCAGACAAAGTTCAATGAATCT 451
807 LeuGlnValAsnThrThrAlaLeuLeuProLysGluAlaAsnSerGluG 823
452 CTTCAAGTGAATAGTCTGCTCATCCCAAGGAGGAGCCCAACTCTGAGGA 501
823 uValPheLeuPheLysProGluAsnIleThrPheGluAsnGlyThrAspL 840
502 AGTCTTTGTTTAAACCAAGAAATCTTCTTTGAAATGGCACAGATC 551
840 euPheIleAlaIleGlnAlaValAspLysValAspLeuLysSerGluIle 856
552 TTTTCATTTGCTATTCAGGCTGTTGATAGGTCGATCTGAAATCAGAAATA 601
857 SerAsnIleAlaArgValSerLeuPheIleProGlnThrProGlnI 873
602 TCCACATTCGACGAGTATCTTTGTTTATCTCCACAGACTCGCCGAGA 651
873 uThrProSerProAspGluThrSerAlaProCysProAsnIleHisIle 889
652 GACACCTAGTCTGTATGAAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 701
890 AsnSerThrIleProGlyIleHisIleLeuLysIleMetTrpLysTrpI 906
702 AACAGCACCATTCTCGCATTCACATTTTAAATAATATGTGGAAGTGGT 751
906 eGlyGluLeuGlnLeuSerIle 913
752 AGGAGAACTGCAGTTGTCAATA 773

seq_name: /cgn2_6/ptodata/2/ins/PCTUS_COMB.seq:PCT-US95-07289-8

seq_documentation_block:
; Sequence 8, Application PCT/US9507289
; GENERAL INFORMATION:
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: Colon Specific Genes and Proteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; ADDRESS: Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
807 LeuGlnValAsnThrThrAlaLeuIleProLysGluAlaAsnSerGluG1 823
|||||
452 CTTCAGTGAATACACTACGCTCTCATCCAAAGGAAGCAACTCTGAGGA 501
|||||
823 uValPheLeuPheLysProGluAsnIleThrPheGluAsnGlyThrAspL 840
|||||
502 AGTCTTTTGTGTTAAACCAAGAAACATTTACATTTTGAAATGGCAGATC 551
|||||
840 euPheIleAlaIleGlnAlaValAspLysValAspLeuLysSerGluIle 856
|||||
552 TTTTCATTGCTATTCAAGGCTGTGATAAGTCCGATCTGAAATCAGAAATA 601
|||||
857 SerAsnIleAlaArgValSerLeuPheIleProProGlnThrProProG1 873
|||||
602 TCCAACATGTCAGCAGATATCTTTGTTTATTCCTCCACAGACTCCGCCAGA 651
|||||
873 uThrProSerProAspGluThrSerAlaProCys.ProAsnIleHisIle 889
|||||
652 GACACCTAGTCTGATGAACGCTGCTCCTTGCTTAATATTCATATC 701
|||||
890 AsnSerThrIleProGlyIleHisIleLeuLysIleMetTrpLysTrpI1 906
|||||
702 AACAGCACCATTCTGGCATTACATTTTAAATAATATGTGGAAGTGGT 751
|||||
906 eGlyGluLeuGlnLeuSerIle 913
|||||
752 AGGAGAACTGCAGTTGTCAATA 773
|||||
```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-221-298-34

seq_documentation_block:

; Sequence 34, Application US/09221298

; Patent No. 6284241

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS

; FILE REFERENCE: 210121.471

; CURRENT APPLICATION NUMBER: US/09/221,298

; CURRENT FILING DATE: 1998-12-23

; NUMBER OF SEQ ID NOS: 112

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 34

; LENGTH: 401

; TYPE: DNA

; ORGANISM: Human

US-09-221-298-34

alignment_scores:
Quality: 618.00 Length: 136
Ratio: 4.754 Gaps: 6
Percent Similarity: 95.588 Percent Identity: 95.588

alignment_block:
US-09-049-696-41 x US-09-221-298-34

Align seg 1/1 to: US-09-221-298-34 from: 1 to: 401

```
29 AsnAsnGlyTyrGluGlyIleValValAlaIleAspProAsnValProG1 45
|||||
1 AACATGCTATGAGGATTTGCTTCAATCGACCCCAATGTGCCAGA 50
|||||
45 uAspGluThrLeuIleGlnGlnIleLysAspMetValThrGlnAlaSerL 62
|||||
51 AGATGAACACTCATTCACAAATAAAGGACATGTTGACCCAGGCATCTC 100
|||||
62 euTyrLeuPheGluAlaThrGlyLysArgPheTyrPheLysAsnValAla 78
|||||
101 TGTATCTGTTGAAGCTACAGGAAGCGATTTTATTTCAAAAAATGTTGCC 150
|||||
79 IleLeuIleProGluThrTrpLysThrLysAlaAspTyrValArgProLy 95
|||||
```

```
151 ATTTTGATTCTCTGAACATGGAACACAAAGGCTGACTATGTGAGACCAA 200
|||||
95 sLeuGluThrTyrLysAsnAlaAspValLeuValAla.GluSerThrPro 111
|||||
201 ACTTGAGACCTACAAAAATGCTGATGTTCTGTTGCTTGAGCTACTCCT 250
|||||
112 ProGlyAsnAspGluProTyrThrGluGlnMet.GlyAsnCysGlyGluL 128
|||||
251 CCAGGTAATGATGAACCCCTACACTGAGCAGATGGGCAACTGTGGAGAGA 300
|||||
128 ys.GlyGluArgIle.HisLeuThrProAspPheIleAlaGlyLysLysL 144
|||||
303 AGGGGTGAAGGATCCACACTCACTCTCGATTTCATTCGAGGAAAAAGT 350
|||||
144 euAla.GluTyrGlyProGlnGly.ArgAlaPheValHisGluTrp 158
|||||
351 TAGCTTGAATATGACCACCAAGGTAAAGGCATTTGTCCATCAATGG 396
|||||
```

seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-09-385-982-23

seq_documentation_block:

; Sequence 23, Application US/09385982

; Patent No. 6262334

; GENERAL INFORMATION:

; APPLICANT: ENDEGE, WILSON O., ET AL.

; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

; FILE REFERENCE: CCDNA-260XX

; CURRENT APPLICATION NUMBER: US/09/385,982

; CURRENT FILING DATE: 1999-08-30

; EARLIER APPLICATION NUMBER: 09/328,111

; EARLIER FILING DATE: 1999-06-08

; EARLIER APPLICATION NUMBER: 60/117,393

; EARLIER FILING DATE: 1999-01-27

; EARLIER APPLICATION NUMBER: 60/098,639

; NUMBER OF SEQ ID NOS: 544

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 23

; LENGTH: 576

; TYPE: DNA

; ORGANISM: Homo sapiens

; NAME/KEY: misc_feature

; LOCATION: (1)...(576)

; OTHER INFORMATION: n = A,T,C or G

US-09-385-982-23

alignment_scores:
Quality: 519.00 Length: 181
Ratio: 3.629 Gaps: 2
Percent Similarity: 79.006 Percent Identity: 60.221

alignment_block:
US-09-049-696-41 x US-09-385-982-23

Align seg 1/1 to: US-09-385-982-23 from: 1 to: 576

```
620 ProfileLeuArgAlaSerValThrAlaLeuIleGluSerValAsnGlyLy 636
|||||
2 CCTGTTCTTGAGCCCAATGCTGACTGCTTCATTGAATCAGAGATGGACA 51
|||||
636 sThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAlaAspAlaThrL 653
|||||
52 TACAGAAGTTTTTGAACCTTTTGGATAATGTTGCGAGCGCTGATCTTTCA 101
|||||
653 ysAspAspGlyValTyrSerArgTyrPheThrThrTyrAspThrAsnGly 669
|||||
102 AGAATGATGGAGTCTACTCCAGGTATTTTACAGCATATACAGAAATGCG 151
|||||
670 ArgTyrSerValLysValArgAlaLeuGlyGlyValAsnAlaAlaArgA 686
|||||
```

```
152 AGATATAGCTTAAAGTTCGGGCTCATGGAGGAGCAAAACACTCCAGGCT 201
686 gArGValIleProGlnGlnSerGlyAlaLeuTrpIleProGlyTrpIleG 703
202 AAAATTACGGCTCCACTGAATAGACCGCGTACATACCAGCTGGGTAG 251
703 luAsnAspGluIleGlnTrpAsnProProArgProGluIleAsnLysAsp 719
252 TGAACGGGGAATTAAGCAAAACCCCAAGACCTGAATTGAT...GAG 298
720 AspValGlnHisLysGlnValCysPheSerArgTrpSerSerGlyClySe 736
299 GATACCTCAGACACCTTGGAGGATTTTCAGCCGAACCATCCGGAGGTGC 348
736 rPheValAlaSerAspValProAsnAlaProIleProAspLeuPheProP 753
349 ATTGTGGTATCAGAGTCCAGCCCTCTTG...CCTGACCAATACCCAC 397
753 roGlyGlnIleThrAspLeuLysAlaGluIleHisGlyGlySerLeuIle 769
398 CAAGTCAATACAGACCTTGATGCCACAGTTTCATTAGGATAGATTATT 447
770 AsnLeuThrTrpThrAlaProGlyAspTrpAspHisGlyThrAlaHi 786
448 ...CTTACATGGACAGCACAGAGATAATTTTGATGTTGGAAGAGTTCA 494
786 sLysTrpIleLeuArgIleSerThrSerIleLeuAspLeuArg 800
495 ACGTTATATCATAGATAATGCGCAGTATCTTGACTAAGAGA 537
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-385-982-25

```
seq_documentation_block:
; Sequence 25, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25
; LENGTH: 595
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(595)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-25
```

alignment_scores:

Quality	468.50	Length:	172
Ratio:	3.496	Gaps:	3
Percent Similarity:	77.907	Percent Identity:	56.395

alignment_block:

US-09-049-696-41 x US-09-385-982-25 ..

Align seg 1/1 to: US-09-385-982-25 from: 1 to: 595

698 ileProGlyTrpIleGluAsnAspGluIleTrpAsnProProArgPr 714
||||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||

```
3 ATACACAGCTGGTAGTGAACGGGAAATTGAACAAACCCGCCAAGACC 52
714 oGluIleAsnLysAspValGlnHisLysGlnValCysPheSerArgT 731
|||||: : : : : |||||
53 TGAATTGAT...GAGGATAGTCAGACCACCTTGGAGATTTCAGCCGAA 99
731 hrSerSerGlyGlySerPheValAlaSerAspValProAsnAlaProIle 747
|||||: : : : : |||||: : : : : |||||: : : : : |||||:
100 CAGCATCCGGAGGTGCATTTGTGTTNTCACAAGTCCCAAGCCTTCCCTTG 149
748 ProAspLeuPheProGlyGlnIleThrAspLeuLysAlaGluIleHi 764
|||||: : : : : |||||: : : : : |||||: : : : : |||||
150 CCTGACCAATACCCACCAAGTCAATACACACCTTGATGCCACAGNTCA 199
764 sGlyGlySerLeuLeuAsnLeuThrTrpAlaProGlyAspTrpA 781
|||||: : : : : |||||: : : : : |||||: : : : : |||||:
200 TGAGGATAAATTATT...CTTACATGGACAGCACCCAGGAGATAATTG 246
781 sPHisGlyThrAlaHisLysTrpIleIleArgIleSerThrSerIleLeu 797
|||||: : : : : |||||: : : : : |||||: : : : : |||||
247 AGTTGGAAAGTTTCACGNTATATCATAGAATAAAGTGCAGATTCTT 296
798 AspLeuArgAspLysPheAsnGluSerLeuGlnValAsnThrThrAlaLe 814
|||||: : : : : |||||: : : : : |||||: : : : : |||||
297 GATCTAAGAGACAGTTNTGATGATGCTCTTCAAGTAAATACTACTGATCT 346
814 uIleProLysGluAlaAsnSerGluGluValPheLeuPheLysProGluA 831
|||||: : : : : |||||: : : : : |||||: : : : : |||||
347 GTCACCAAGGAGGCCAACTCCAANGAAGCTTTCGNTTTAAACAGAAA 396
831 snIleThrPheGluAsnGlyThrAspPheIleAlaIleGlnAlaVal 847
|||||: : : : : |||||: : : : : |||||: : : : : |||||
397 ATATCTCAGAGAATAATGCAACCCACATATTTATTCGCCNTTNAAGTATA 446
848 AspLysValAspLeuLysSerGluIleSerAsnIleAlaArgValSerIe 864
||| : : : : : |||||: : : : : |||||: : : : : |||||
447 GATANAGCA...ATTGACATCNAAGTNTCCACATTGNACAAAGTNACTTT 493
864 uPheIleProGln 869
493 GGTATFCCCTCAGCAA 509
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-385-982-24

```
seq_documentation_block:
; Sequence 24, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 618
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(618)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-24
```



```

CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/098,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 544
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 27
LENGTH: 611
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(611)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-27

alignment_scores:
Quality: 421.50 Length: 177
Ratio: 3.345 Gaps: 5
Percent Similarity: 71.186 Percent Identity: 52.542

alignment_block:
US-09-049-696-41.x US-09-385-982-27 ..
Align (seg 1/1 to: US-09-385-982-27 from: 1 to: 611

630 ProLeuAlaSerValThrAlaLeuIleGluSerValAsn.GlyL 636
||||| ||||| ||||| ||||| ||||| ||||| |||||
2 CCGTCTCTGGAGCCAAATGTGACTGCTTTCATTGAATCAGAAATGGAC 51
636 ysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAlaAspAlaThr 652
||||| ||||| ||||| ||||| ||||| ||||| |||||
52 ATACAGAGCTTTTGGAACTTTTGGATAATGGTCAGCGGCTGATTCTTC 101
653 LysAspAspGlyValTyrSerArgTyrPheThrThrTyrAspThrAsnGl 669
||||| ||||| ||||| ||||| ||||| ||||| |||||
102 AAGATGATGGAGCTCTACTCCAGGTATTTTACAGCATATACAGAAATGG 151
669 yArgTyrSerValLysValArgAlaLeuGlyGlyValAsnAlaAlaArgA 686
||||| ||||| ||||| ||||| ||||| ||||| |||||
152 CAGATATAGCTTAAAGATTGGGCTCATGGAGGAGCAACACTGCCAGGC 201
686 rGArgValIleProGlnGlnSerGlyAlaLeuTyrIleProGlyTrpIle 702
||||| ||||| ||||| ||||| ||||| ||||| |||||
202 TAAATATTACGGCTCCACTGAGTAGAGCCGCTACATACCAGTGGGTA 251
703 GluAsnAspGluIleGlnTrpAsnProProArgProGluIleAsnLysAs 719
||||| ||||| ||||| ||||| ||||| ||||| |||||
252 GTCAACGGGGAATTCAGCAACACCCGCCAAGACCTGAAATTGAT...GA 298
719 pAspValGlnHisLysGlnValCysPheSerArgThrSerSerGlyGlyS 736
||||| ||||| ||||| ||||| ||||| ||||| |||||
299 GGATACTCAGACCACTTGGAGGATTTTCAGCCGAACAGCATCCCGAGGTG 348
736 erPheValAlaSerAsp.ValProAsnAlaProIleProAspLeuPhePr 752
||||| ||||| ||||| ||||| ||||| ||||| |||||
349 CATTTGTGATACAAAGTCCCAACACTTTCCTTGCCTGACCAATACCC 398
752 oProGlyGlnIleThrAspLeuLysAlaGlu..... 762
||||| ||||| ||||| ||||| ||||| ||||| |||||
399 ACCAAGTCAAAATCACAGACCTTGATGCCACCAAGTCATTAGGATAAAATP 448
763 ..IleHisGlyGlySerLeuIleAsnLeuThrTrpThrAlaProGlyLys 778
||||| ||||| ||||| ||||| ||||| ||||| |||||
449 TCTTACATCGA.....NGCCCAANGAAT 471
779 AspTyrAspHisGlyThrAlaHisLys 787
||||| ||||| ||||| ||||| ||||| ||||| |||||
472 AATTTGTGATTGNGNAAGNACCTACCGT 498

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-385-982-33

seq_documentation_block:
; Sequence 33, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCNA-260XX
; CURRENT FILING DATE: 1999-08-30
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33
; LENGTH: 742
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(742)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-33

alignment_scores:
Quality: 373.50 Length: 131
Ratio: 3.557 Gaps: 4
Percent Similarity: 80.153 Percent Identity: 58.779

alignment_block:
US-09-049-696-41 x US-09-385-982-33

Align seg 1/1 to: US-09-385-982-33 from: 1 to: 742

698 ILeProGlyTrpIleGluAsnAspGluLeuTrpAsnProArgPr 714
|||||
3 ATACAGGCTGGTAGTGAAGGGGAATTGAAGCAACCGCCAGACC 52
714 oGluLeuAsnLysAspValGlnHisLysGlnValCysPheSerArgT 731
|||||
53 TGAATTGAT...GAGGATACCTAGACACCCTTGAGGATTTTCAGCGGAA 99
731 hrSerSerGlySerPheValAlaSerAspValProAsnAlaProIle 747
|||||
100 CACATCCGGAGGTGATTTGGTATCAAGTCCCAAGCCTCCCTTG 149
748 ProAspLeuPheProGlyGlnIleThrAspLeuLysAlaGluIleH 764
|||||
150 CCGACCAATACCCACCAAGTCAATCAATCAGACCTTGATGCCACAGTCA 199
764 sGlyGlySerLeuIleAsnLeuThrTrpAlaProGlyAspAspTyrA 781
|||||
200 TGAGGATAAGATTATT...CTTACATGGACAGCAGCAGGAGATAATTG 246
781 sphisGlyThrAlaHisLysTyrIleIleArgIleSerThrIleLeu 797
|||||
247 ATCTTGAAAAAGTTCAACGTATATCATAGATAAGTCAAGTATCTT 296
798 AspLeuArgAspLysPheAsnGluSerLeuGlnValAsnThrThrAlaLe 814
|||||
297 GATCTAAGACACAGATTGATGATGCTCTTCAAGTAAAT...CTACTGAT 343
814 u.IleProLysGluAlaAsnSerGluGluValPheLeuphe 827
|||||
344 CTCACCAAGGAGGCGCACTTCAAGGAAGCTTTGCAATT 384

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-617-697-8

seq_documentation_block:
; Sequence 8, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; FILE REFERENCE: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstreser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4803 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-617-697-8

alignment_scores:
Quality: 139.50 Length: 684
Ratio: 0.423 Gaps: 30
Percent Similarity: 48.246 Percent Identity: 20.468

alignment_block:

US-09-049-696-41 x US-08-617-697-8

Align seg 1/1 to: US-08-617-697-8 from: 1 to: 4803

281 AspPheLysLysThrThrProMetThrGlnProProAsnProThrPh 297
|||
255 GACGCCAATAAAACCACTATCCGTAATAGCTCATCATCATCATG 305
297 eSerLeuLeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLys 314
|||
306 GAAACAATTAACATTGACCAAAATGAAATGGTGCAGTTTTCACAGAAA 355
314 erGlySerMetAlaThrGlyAsnArgLeuAsnArg..... 325
|||
356 GCAGCAACTCTCGCGTTTCAACCGGTGTACATCTGCACCAAAATCTCCCAA 405
326LeuAsnGlnAlaGlyGlnLeuPheLeuLeuGln..... 336

[illegible]

1273	TCCATTGGTGATGATGATGATGTTGTCACGCTAAAGAGTGGTTATTAGACCC	1322
578ThrValThrSerArgAlaSerAsnAlaThrL	588
1323	AGATGATGTGCCATTGAAACTTACATCTGGACGCAATAATACCGCG	1372
586	euProProlleThrValThrSerLysThrAsnLysAspThrSerLys...	603
1373	AAAACCAAGGATATACAAACAGGAGATGGGACTTAAGAGTCACTAAAGT	1422
604PheProSerProLeuValValThrAlaAsnIleArgGlnGlyAl	618
1423	AATAGTATTCTTAAACCTACATTAACAACTCAACTCTTGAGCAA.....	1467
618	asrProIleLeuArg.....AlaSerValThrAla.....	628
1468ATCCTAAGAGAGGTTCTTATGTTAATATCACTGCTAATAATA	1510
629LeuIleGluSerValAsnGlyLysThrValThrLeuGluLeu	642
1511	GAATTTATGTTAATAGTCCATCAAC.....	1536
643	LeuAspAsnGlyAlaGlyAlaAspAlaThrLysAspAspGlyValTyrSe	659
1537	TTATCATTAAGGAGTTTAACACTTCACACTAACAGAGATGGATTAATAAT	1586
659	rArgTyrPheThrThrTyrAspThrAsnGlyArgTyrSerValLysVala	676
1587	TACGGTGATATTACCTCAACGAAATGGTAATTAACCATTAAGCAG	1636
676	rgAlaLeuGlyGlyValAsnAlaAlaAraGArGArgValIleProGlnGln	692
1637	GCTCTGGTGGTGGTTCAT.....AAAAACATCAGCTTGGT	1674
693	SerGlyAlaLeuTyrIleProGlyTrpIleGluAsnAspGluIleGlnTr	709
1675	ACGGGTTTTTGAATATT.....CTCGCTGGGAGTCTGTAGCTTT	1715
709	pAsnProProArgProGlu.....	715
1716	TGAGAGAGGGCGATAAAGCAGTAAACGCAACAGATGCTCAANTACCG	1765
716IleAsnLysAspAspValGlnHisLys.....	724
1766	CACAAGGGACGATAACCGTCAATAAAGATGATAACAATTTAGATTCAAT	1815
725	GlnValCysPheSerArgThrSerSerGlyGlySerPheValAlaSerAs	741
1816	AATGTATCTATTAAACGGGACGGGCAAGGGTTAAAGTTATTGCAATCA	1865
741	pValProAsnAlaProlleProAspLeuPheProGlyGlnIleThr.	757
1866	A.....ATAATTTTCACCTC	1879
758AspLeuLysAlaGluIleHisGlyGlySerLeuIleAsnLeu	771
1880	ATAAATTGTATGGCAAAATTAACATATCTGGAATAGTAACAATTAACCAA	1929
772	Thr.....TrpThrAlaProGlyAspAspTyrAs	781
1930	ACCACGAAAAAAGATGTAAATACCTGGAATGCATCAAAAGACTCTTACTG	1979
781	pHis.....GlyThrAlaHisLysTyr.....	788
1980	GAATGTTCTCTTACTTTGAATACGGTGCMAAAATTTACCTTTATAA	2029
789IleIleArgIleSerThrIleLeuAspLeuArgAspLysPheAsn	804
2030	AATTCGTTGATACGGGCTCAAAATCCCAAGATTGAGGTATCAGCTAGA	2079
805	GluSerLeuGlnValAsnThrThrAlaLeuIleProLysGlu.....	818
2080	AGTTTTGCAGCGGTACATTTTAAACGGCATTCGAGGCAAAACAACTTCAA	2129

407 lIeValLeuLeuThrAspGlyCluAspAsnThrIleSerGlyCysPheAs 423
2116GGTAAAGTAACTGTAAAGAGATGCTACGCTAA 2147
423 nGluValLysGlnSerGlyAlaIleIleHisThrValAlaLeuGlyProS 440
IleHisThrValAlaLeuGlyProS 440
2148 TGACGCTGTAAGAAGTCGCAACTGTAAAGATGTTGCACCGCAATTA 2197
440 eAlaAlaGlnLeuGluLeuSerLysMetThrGlyGlyLeuGln 456
IleHisThrValAlaLeuGlyProS 440
2198 ATAGTGCAGCGCTTTTGTAAACAGAGAAATTAACCTACCTCTATTGAT 2247
457 ThrTyrAlaSerAspGlnValGlnAsnGlyLeuIleAspAlaPheCl 473
2248GAAGATAATCCTACAGATACGCGCAAGATGAC..... 2280
473 yAlaLeuSerSerGlyAsn.....GlyAlaValSerGlnArgS 486
IleHisThrValAlaLeuGlyProS 440
2281 GCACCTTAAGCGGCGATACCTTAACCTTTAAACGAGGTAAACCTGA 2329
486 eTlleGlnLeuSerLysGlyLeuThrLeu.....GlnAsnSer 499
IleHisThrValAlaLeuGlyProS 440
2330 AAGTTAAACGCTGATGCAAAATATATCTTTTACTTGGCNAACCTT 2379
500 GlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspTh 516
2380 GAGGTGAACCTGCGAAGAGTACTTAAACGATTGGCGGAATAC 2429
516 rLeuPheLeuIleThrTrpThrThrGlnProGlnIle..... 529
2430 A.....CCTACAGTGGCCTACTTACGACGCAAAAGTATATACTA 2473
529 529
2474 GCACGGCTGATGTTGAATTTTCAAAAGAAACAGCGGATGCCCTCGGT 2523
530LeuLeuTrpAspProSe 535
2524 TCTAAGATGTTTATTGAAGGTATTCGCAACTTTAACTGAGCCCAAG 2573
535 rGlyGlnLysGlnGlyPheVal.....ValAspLysAsnThrL 549
2574 CGCGGAGCGAAGTCTTACACGTTGATTAAATGTGGATGGACGAGAAA 2623
549 ysmetAlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLys 565
2624 AATCAATGCCAGCAAGTATTGAAGATGATTGCGCGCAGGT..... 2664
566 TyrSerLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSerAr 582
2665 TGGATATTCAGGTAAGTAAATGTTGATTTATGAGCAGCATATCA 2714
582 gAlaSerAsnAlaThr.....LeuProProlIleThrValT 594
2715 CACAGTAACTTTACCGATGACAGCAGGTACAAACAGGTAAACCGTAA 2764
594 hrSerLysThrAsnLysAspThrSerLysPheProSerProLeuValVal 610
2765 CCAAAAAAGCAGATGGCAAGGT..... 2787
611 TyrAlaAsnIleArgGlnGlyAlaSerProIleLeuArgAlaSerValTh 627
2788 ...GCTGAGTTAAATCGGTGCGAAA.....AC 2813
627 rAlaLeuIleGluSerValAsnGlyLysThrValThr...LeuGluLeuL 643
2814 TTCTGTTATCAAGACCAACAGGCAACTGTTTACAGCAAGACCTGA 2863
643 euAspAsnGlyAlaGlyAlaAspAlaThrLysAspGly..... 656
2864 AAGATCGCAATTAATGTCACCGTGTAGTGAAGATGATGCGCAAGACACC 2913
657ValTyrSerArgTyrPheThrThrThrAspThrAsnG 669

2914 GCACAGCGCTTAGTTACTCAAAA.....ACTGTGATTGATGCAGTAA 2957
669 yArgTyrSerValLysValArgAlaLeuGly..... 679
2958 TAAAGCGGTGGAGGTAAACCGGTGAGGCGCGGACTGCGGAACCGGTG 3007
680GlyValAsnAlaAlaArgArgValIleProGlnGlnSerGly 694
3008 CAACCGCGTGAATCGGTGATCGGTGAAACCGTTACATCAGCAGCAGC 3057
695 AlaLeuTyrIleProGlyTrpIleGluAsnAspGluIleGlnTrpAsnPr 711
3058 GTGAACCTTCAAAAACGCG.....AATGC 3080
711 oProArgProGluIleAsnLysAsp.....A 720
3081 GACCACAGCGACCGTAAGCAAGATAATGCAACATCAATGTCAAAATACG 3130
720 spVal.....GlnHisLysGlnValCys 727
3131 ATGTAATGTTGTCACGGCTTGAAGATTGGCGATGACAAAAAATCGTT 3180
728 PheSerArgThrSer.....SerGlyGlySerPheVal..... 738
3181 GCAGACACGACACACTTACTGTAAACGGTGGTAAGGTGCTGCTCTCTGC 3230
739AlaSerAspValProAsnAlaProIleProAspLeuPheProG 754
3231 TGGTCTAATAGTTAATAAC..... 3252
754 lyGlnIleThrAspLeuLysAlaGluIleHisGlySerLeuIleAsn 770
3253 ..AATAAGAACTTGTAAATGAGAGGTTTACGCGACTGCTTTAAACAAC 3300
771 LeuThrTrpThrAlaProGlyAspTyrAspHisGly 783
3301 CTAAGCTGGACGCAAAAGCCGATAAATATGCGATGGC 3339
seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-08-685-467-3
seq_documentation_block:
Sequence 3, Application US/08685467
Patent No. 6060059
GENERAL INFORMATION:
APPLICANT: St. Gene III, Joseph W.
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Honbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 22-JUL-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
TELECOMMUNICATION INFORMATION:

1496 TTTCAAAAGACCAAGAAGTCAAAAGCAGCGGAA..... 1527
134 LeuThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGlu 150
1528 ...ACGGTAACCTTTAAAGCGGCAAGAACTTA...AAAGTGAACAGGA 1571
150 nGlyArgAlaPheValHisGlu.....TrpAlaH 160
1572 TGGTGCAACTTTACITATTCACTGCAAGATGCTTTAAACGGGTTTAAACGA 1621
160 isLeuArgTyrGlyValPheAspGluTyrAsnAsnAspGluLysPheTyr 176
1622 GCATTACTTTAGTGTGTACAACTAATAGCGGAAATGATGCGAAACCGTC 1671
177 LeuSer...AsnGlyArgIleGlnAlaValArgCysSerAlaGlyIleH 192
1672 ATCAACAACACCGGTTTAAACCATCACCACAGAGTAAAGTAAAGCAGGTAATA 1771
192 rGlyThrAsnValValLysLysCysGlnGlyGlySerCysTyrThrLysA 209
1722 AGGTACAAACACCATCAGCGTAACCAAGATGGCATTAAAGCAGGTAATA 1771
209 rGlyThrPheAsnLysValThrGlyLeu.....TyrGluLysGlyCys 223
1772 AAGCTATTACTAATGTTGCGAGTGTGTTAAGAGCTTATGACGATGCGAAT 1821
224 GluPheValLeuGlnSerArgGlnThrGluLysAlaSerIleMetPheAl 240
1822 TTTGATGTTTAAATAACTCTGCAACTGAT.....TTAAA 1856
240 aGlnHisValAspSerIleValGluPheCysThrGluGlnAsnHisAsnL 257
1857 TAGACACGTTCAAGATGCTTATAAAGGTTTATTA.....AATCTAAATG 1900
257 ysGluAlaProAsnLysGlnAsnGlnLysCysAsnLeuArgSerThrTyr 273
1901 AAAAATGCAATTAACAA..... 1920
274 GluValIleArgAspSerGluAspPheLysLysThrThrProMetThrTh 290
1921CGTTTGGTGCAC 1931
290 rGlnProAsnProThrPheSerLeuLeuGlnIleGlyGlnArgIleV 307
1932 TGACAGCAGCGGCGACTGTAGCGGATTTA.....CGTAAATGG 1972
307 aLysLeuValLeuAspLysSerGlySerMetAlaThrGlyAsnArgLeu 323
1973 GTTGGGTAGTATCAACCAAAACGCT.....ACGAAAGAAAGAAC 2013
324 AsnArgLeuAsnGlnAlaGlyGlnLeuPheLeuGlnThrValGluLe 340
2014 AATCAAGTTAAACAAGCTGATGAAGTC.....CT 2042
340 uGlySerTrpValGlyMetValThrPheAspSerAlaAlaHisValGln 357
2043 CTTTACCAGCGCGTGTGTCTACGTTACTTCCAAATCTGAAACCGTA 2092
357 erGluLeuIleGlnLysSerGly..... 365
2093 AACATACGATTACCGTTAGTGTGCTGAACTAAAGCGGATTCGGTCTT 2142
366 ...SerAspArgAspThrLeuAlaLysArgLeuPro..... 376
2143 GAAAGAGTGGCGATCTATTAAAGCTCAAGTGGATATCAAAACACATCA 2192
377AlaAlaSerGlyGlyThrSerIleCysSerGlyLeuA 390
2193 TAATGTTTAACTGTTGGTAAATAGTGTCTGCTCTCACTAAAGGT... 2238
390 rGlySerAlaPheThrValIleArgLysLysTyrProThrAspGlySerGlu 406
2239GGCTTTGAAGTGT...AAACGAGGAGGAGTGTGATGATGATG 2280

407 IleValLeuLeuThrAspGlyGluAspAsnThrIleSerGlyCysPheAs 423
2281GGTAAAGTAACTGTAAAGATGCTACTGCTAA 2312
423 nGluValLysGlnSerGlyAlaIleHisThrValAlaLeuGlyPro 440
2313 TGACCTGATATAAGAAAGTCGCAACTGTAAGATGTTGCAACCGCAATA 2362
440 erAlaAlaGlnGluLeuGluLeuSerLysMetThrGlyGlyLeuGln 456
2363 ATAGTGGCGCGACTTTGTGAAACACAGAGATTTAACTACCTCTATTGAT 2412
457 ThrTyrAlaSerAspGlnValGlnAsnGlyLeuIleAspAlaPheGlu 473
2413GAAGATAATCCTACAGATAACGCGCAAGATGAC..... 2445
473 yAlaLeuSerSerGlyAsn.....GlyAlaValSerGlnArg 486
2445 GCACCTTAAAGCGGCGATACCTTAACCTTTAAAGCAGGTAAACACCTGA 2494
485 erIleGlnLeuGluSerLysGlyLeuThrLeu.....GlnAsnSer 499
2495 AAGTTAAACGTGATGGAATAATATTTGACTTGGCGAAACACCTT 2544
509 GlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspTh 516
2545 GAGGTGAAACACTGCGAAAGTGAGTACTTTAAACGATTTGGCGGGAATAC 2594
516 rLeuPheLeuIleThrTrpThrGlnProGlnIle..... 529
2595 A.....CCTACAGGTGGCAGTACTCGCAGCGCAAAAGTGAATATTACTA 2638
529 529
2639 GCACGCTGATGCTTTGAAATTTTGCAAAAGAAACACCGATGCTCGGT 2688
530LeuLeuTrpAspProse 535
2689 TCFAAGAAATGTTTATTGAAAGTATTGCGCAACATTTAACTGAGCGCAAG 2738
535 rGlyGlnLysGlnGlyPheVal.....ValAspLysAsnThrL 549
2739 CGCGGGAGCGAAGCTTTCACACGTTGATTTAAATGTGGTGCAGCAAAA 2788
549 yMetAlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLys 565
2789 AATCCAAATGCAAGATTTGAAGATGATTTGCGCGAGGT..... 2829
566 TyrSerLeuGlnAlaSerGlnThrLeuThrLeuThrValThrSerAr 582
2830 TGGAAATTCAGGTAAATGTAATATGTTGATTATGATGACGCTATGA 2879
582 gAlaSerAsnAlaThr.....LeuProIleThrVal 594
2880 CACAGTAAACTTTACCGATGACAGCAGCAGTACACACCGTAACCGTAA 2929
594 hrSerLysThrAsnLysAspThrSerLysPheProSerProLeuValVal 610
2930 CCAAAAGCAGATGGCAAGGT..... 2952
611 TyrAlaAsnIleArgGlnGlyAlaSerProIleLeuArgAlaSerValTh 627
2953 ...GCTGACGTTAAATCGGTGCAAA.....AC 2978
627 rAlaLeuIleGluSerValAsnGlyLysThrValThr...LeuGluLeu 643
2979 TTTGTTATCAAGACCAACACGCAACTGTTTACAGGCAAGACCTGA 3028
643 euAspAsnGlyAlaGlyAlaAspAlaThrLysAspAspGly..... 656
3029 AAGATGCAATAATGTCACACCGTTAGTGAAGATGATGCGCAAGACACC 3078

657ValTyrSerArgTyrPheThrThrTyrAspThrAsnG1 669
3079 GGCACAGCTTAGTTACTGAAAA.....ACTGTGATTGTCAGTAAA 3122
669 yArgTyrSerValIysValArgAlaLeuGly..... 679
3123 TAAAGCGGTTGGAGGTAAACCGGTGAGGCGCGACTCCGGAACCGGTG 3172
680GlyValAsnAlaAlaArgArgValIleProGlnGlnSerGly 694
3173 CAACGCGCGTGAATCGGCTAACGCTGAACCGTTACATCAGGCAGCAGC 3222
695 AlaLeuTyrIleProGlyTyrIleGluAsnAspGluIleGlnTyrAsnPr 711
3223 GTGAACITCAAAACGGC.....AATGC 3245
711 oProArgProGluIleAsnLysAsp.....A 720
3246 GACCACAGCAGCGTAAAGCAAAAGATAATGGCAACATCAATGTCAAATACG 3295
720 spVal.....GlnHisLysGlnValCys 727
3296 ATGTAATGTTGGTGACGCTTGAAGATTGGCGATGACAAAAAATCGTT 3345
728 PheSerArgThrSer.....SerGlyGlySerPheVal..... 738
3346 GCAGACAGCAGCACACTACTACTGACAGCTGGTGAAGGTGCTGTTCCGTCG 3395
739AlaSerAspValProAsnAlaProIleProAspLeuPheProG 754
3396 TGGTGCTAATAGTGTAAATAC..... 3417
754 lyGlnIleThrAspLeuLysAlaGluIleHisGlyGlySerLeuIleAsn 770
3418 ..AATAAGAACTTGTGTAATGACAGAGGTTAGCGACTGCTTTAAACAAC 3465
771 LeuThrTyrAlaProGlyAspTyrAspHisGly 783
3466 CTAGCTGCGCGGCAAGCGCATAAATATGCAGTGGC 3504

seq_name: /cgn2_6/ptodata/2/ina/5A_comb.seq:US-08-038-682-8

seq_documentation_block:
; Sequence 8: Application US/08038682

; Patent No. 5549897

; GENERAL INFORMATION:

; APPLICANT: BARENKAMP, STEPHEN J

; APPLICANT: ST. GENE III, JOSEPH W

; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Mattare, Ltd

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; STREET: Bldg 1

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202-0286

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/038,682

; FILING DATE: 16-MAR-1993

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: BERKSTRESSER, JERRY W

; REGISTRATION NUMBER: 22,651

; REFERENCE/DOCKET NUMBER: 1038-293

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 415-0810

; TELEFAX: (703) 415-0813

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4702 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-038-682-8

alignment_scores:

Quality: 131.50 Length: 684

Ratio: 0.398 Gaps: 30

Percent Similarity: 48.246 Percent Identity: 20.322

alignment_block:

US-09-049-696-41 x US-08-038-682-8

Align#seg 1/1 to: US-08-038-682-8 from: 1 to: 4702

261 AspPheLysLysThrThrProMetThrThrGlnProProAsnProThrPh 297

41 GACGCCATAAAACCACTATCCGTAATAGCATCAATGCTATCATCAATG 90

297 eSerLeuLeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLys 314

91 GAAACAATTTAACATTTGACCAAAATGAAATGAGCAGTTTTTACAAGAA 140

314 erGlySerMetAlaThrGlyAsnArgLeuAsnArg..... 325

141 GCAGCAACTCTGCCGTTTTCAACCGTGTACATCTGACCAAAATCTCCCAA 190

326LeuAsnGlnAlaGlyGlnLeuPheLeuLeuGln..... 336

191 TTAAGAGGATTTAGATTCTAACGACCAAGCTTTTTTAATCAACCCAAA 240

337ThrValGluLeuGlySerTrpValGlyMetValThrPheAsp. 350

241 TGGTATCAATATGATAAGACGCAATTTATTAACACTAATGCTTTACTG 290

351 ..SerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySer 366

291 CTCTCAGCTAGACATTTCTAACGAAAAACATCAAGCGCGTAAATTCACC 340

367 AspArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGly 383

341 CTGAGCAAAACCAAGATAAAGCACTCGCTGAAATCGTCAATCAGGTTT 390

383 rSerIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLys 400

391 A...ATTACCGTTGGTAAAGACGCTAGCTAAACCTTATTTGGTGCAG 437

400 yrProThrAsp.....GlySerGluIleValLeuLeu 410

438 TGAACACGAGGCGGTGATTAGCGTAAATGCGGTAGTATTCTTTACTT 487

411 ThrAspGlyGluAspAsnThrIleSerGlyCysPheAsnGluValLysG 427

488 GCA...GGCAAAAAATCACCATCAGCGATATAATAATCCA..... 526

427 nSerGlyAlaIleIleHisThrValAlaLeuGlyProSerAlaAlaGln 444

527ACCATCACTTACAGCATTGTCACCTGCAAAACGAGCGATCA 569

444 luLeuGluGluLeuSerLysMetThrGlyGlyLeuGlnThrTyrAlaSer 460

570 ATCTGGCGGATATTTTGCACAAAGGTGTACATTAAATGTCGCCGCTGCC 619

461 AspGlnValGlnAsnGlyLeuIle.....AspAlaPheGlyAlaLe 475

620 ...ACTATTCCCAATAAAGGTAAACTTCTGCCGACTCTGTAAAGCAAGA 666

709 pAsnProArgProGlu. 715
 1501 TGAGAGAGGGCGATAAAGCAACGCTACGCACAGATGCTCAAAATTACCG 1550
 716IleAsnLysAspValGlnHisLys. 724
 1551 CACAAGGACCATAAACGCTCAATAAGATGATCAAAACATTTAGATTCAAT 1600
 725 GlnValCysPheSerArgThrSerSerGlyCysSerPheValAlaSerAs 741
 1601 AATGATCTATTACGGAGCGGACGGTTAAAGTTTATTCGNAATCA 1650
 741 pValProAsnAlaProIleProAspLeuPheProGlyGlnIleThr. 757
 1651 A.....AATAATTTCACT 1664
 758AspLeuLysAlaGluIleHisGlyGlySerLeuIleAsnLeu 771
 1665 ATAATTTGATGGCGAATTAACATATCTCGAATAGTAAACAATTAAACAA 1714
 772 Thr.....TrpThrAlaProGlyAspAspTyrAs 781
 1715 ACCACGAAAAAGATGTTAAATACTGGAACTCATCAAAAGACTCTTACTG 1764
 781 PHis.....GlyThrAlaHisLysTyr..... 788
 1765 GAATGTTTCTCTCTACTTTGAAATACGGTCGCAAAATTTACCTTTATAA 1814
 789IleIleArgIleSerThrSerIleLeuAspLeuArgAspLysPheAsn 804
 1815 AATTCGTTGATAGCGGCTCAAAATCCCAAGATTTGAGGTCAATCAGTAGA 1864
 805 GluSerLeuGlnValAsnThrThrAlaLeuIleProLysGlu..... 818
 1865 AGTTTTCAGGCGTACATTTTAACGGCATCGGAGCAAAACAACCTCAA 1914
 819AlaAsnSerGluGluValPheLeuPheLysProGluAsnIleT 833
 1915 CATCGGAGCTAACGAAAAGCGCTATTTTAAATTTAAACCAACCGCGCTA 1964
 833 hr 833
 1965 CA 1966

```

seq_documentation_block:
: Sequence 8, Application US/08302832
: Patent No. 5603938
: GENERAL INFORMATION:
: APPLICANT: Barekamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/302,832
: FILING DATE: 16-SEP-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1

```

[illegible][illegible]

552	tyrLeuGlnIleProGly.....	557ileA	559
553	558	560
554	559	561
555	560	562
556	561	563
557	562	564
558	563	565
559	564	566
560	565	567
561	566	568
562	567	569
563	568	570
564	569	571
565	570	572
566	571	573
567	572	574
568	573	575
569	574	576
570	575	577
571	576	578
572	577	579
573	578	580
574	579	581
575	580	582
576	581	583
577	582	584
578	583	585
579	584	586
580	585	587
581	586	588
582	587	589
583	588	590
584	589	591
585	590	592
586	591	593
587	592	594
588	593	595
589	594	596
590	595	597
591	596	598
592	597	599
593	598	600
594	599	601
595	600	602
596	601	603
597	602	604
598	603	605
599	604	606
600	605	607
601	606	608
602	607	609
603	608	610
604	609	611
605	610	612
606	611	613
607	612	614
608	613	615
609	614	616
610	615	617
611	616	618
612	617	619
613	618	620
614	619	621
615	620	622
616	621	623
617	622	624
618	623	625
619	624	626
620	625	627
621	626	628
622	627	629
623	628	630
624	629	631
625	630	632
626	631	633
627	632	634
628	633	635
629	634	636
630	635	637
631	636	638
632	637	639
633	638	640
634	639	641
635	640	642
636	641	643

```

alignment_block:
US-09-049-696-41 x US-08-530-198-8  ..
Align seg 1/1 to: US-08-530-198-8 from: 1 to: 4702

281 AspheLysLysThrThrPrometThrThrClnProProAsnProThrPh 297
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
41 GAGCGCAATAAAACCACTATCCGTAATAGATCATCAATGTCATCATCAATG 90

297 eSerLeuLeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLys 314
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
91 GAAACAATTTAACATTCACCAAAATGAATGGAGCAGTTTTTACAAGAAA 140

314 eGlySerMetAlaThrGlyAsnArgLeuAsnArg..... 325
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
141 GCAGCAACTCGCGTTTTCAACCGTGTATCATCTGACCAAAATCTCCCAA 190

325 .....LeuAsnGlnAlaGlyGlnLeuPheLeuLeuGln..... 336
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
191 TTAAAGGGATTTTAGATTCTAACGGACAAGTCTTTTAAATCAACCCAAA 240

337 .....ThrValGluLeuGlySerTrpValGlyMetValThrPheAsp. 350
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 TGGTATCACAAATAGGTAAGACGCAATTTAATACACTAATGCTTTACTG 290

351 ..SerAlaAlaHisValGlnSerGlnLeuIleGlnIleAsnSerGlySer 366
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
291 CTCTACGTAGACATTCTAACGAAACATCAAGGCGCGTAATTTCACC 340

367 AspArgAspThrLeuAlaLysArgLeuProAlaAlaSerGlyGlyTh 383
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
341 CTTGACCAACCAAGGATAAGCACTCGCTCAAATCGGAATCACGGTTT 390

383 rSerIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLys 400
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
391 A...ATTACCGTTGGTAAAGACGGTAGCGTAAACCTTATTGGTGGCAAG 437

400 yrProthrAsp.....GlySerGluIleValLeuLeu 410
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
438 TGAACAACGAGGGCGTGATTAGCGTAATGCGGTAGATTCTTCTTACT 487

411 ThrAspGlyGluAspAsnThrIleSerGlyCysPheAsnGluValysG 427
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

; ZLF: 22202-0280
; COMPUTER READABLE FORM:

```

488 GCA...GGCAAAAAATCACCATCAGCGATATAATAATCA... 526
427 nSerGlyAlaIleHisThrValAlaLeuGlyProSerAlaAlaGlnG 444
527ACCATCACTTACAGCATTCGTCACCTGAAAAACGAAGCATCA 569
444 LuLeuGluLeuSerLysMetThrGlyGlyLeuGlnThrTyrAlaSer 460
570 ATCTGGCGGATATTTTCCAAAGGTGGTAAACATTAAATGTCGCGCTGCC 619
461 AspGlnValGlnAsnGlyLeuLeu.....AspAlaPheGlyAlaLe 475
620 ...ACTATTCCGAATAAAGTAAACTTCTGCCGACTCTGTAAGCAAGA 666
475 uSerSerGlyAsnGlyAlaValSerGlnArgSerIleGlnLeuGluSerL 492
667 TAAAGTGGTAAACATTGTTCTCTGCCAAAGAGGTGAAGCGGAATTG 716
492 ysGly...LeuThrLeuGlnAsnSerGlnTrpMetAsnGlyThrValIle 507
717 GCGGTGTAATTTCCGCTCAAAATCAGCAAGCCAAAGGTGGTAAAGTTGATG 766
508 ValAspSer.....ThrValGlyLysAspThrLeuPheLeuIleTh 521
767 ATTACAGGTGATAAAGTCACATTAAACACAGGTGCAGTTATCGACCTTTC 816
521 rTrpThrThrGlnProProGlnIleLeuLeuTrpAspProSerGlyGlnL 538
817 AGGTAAGAAGGGGAGAGACTTATCTTGGCGGTGATGAGCGTGCGGAAG 866
538 ysGlnGlyGlyPheValValAspLysAsnThr.....LysMetAla 551
867 GTAAAAATGGTATTCAATTAGCGAAGAAACCTCTTTAGAAAAGGCTCG 916
552 TyrLeuGlnIleProGly..... 557
917 ACAATTATGATCAGCGCAAGAAAGCGCGCGCTATTGTTATGGG 966
558 11eA 559
967 CGATATTCGATTAAATATGTTAATGTTAATGCTCAAGGTACGGATATG 1016
559 IaLysValGlyThrTrpLysTyrSerLeuGlnAlaSerSerGlnThrLeu 575
1017 CTAAACTGGCGGCTTT.....GTGGAACATCAGGACATGACTTA 1057
576 ThrLeu..... 577
1058 TCCATTGGTGATGATGATGTTTGCACGCTAAAGAGTGCTTATTAGACC 1107
578ThrValThrSerArgAlaSerAsnAlaThrL 588
1108 AGATGATGTGCTTCAATGAAACTCTTACATCTGGACGCAATATACCGGCG 1157
588 euProProlleThrValThrSerLysThrAsnLysAspThrSerLys... 603
1158 AAACCAAGGATATACACAGGAGATGGGACTAAAGAGTCCTAAAGT 1207
604PheProSerProLeuValValTyrAlaAsnIleArgGlnGlyAl 618
1208 AATAGTATTCTTAAACCTACATTAAACAACTCAACTCTTGACCA... 1252
618 aSerProlleLeuArg.....AlaSerValThrAla..... 628
1253ATCCTAAGAAGAGGTCTTATTATATATCACTGCTAATAATA 1295
629LeuIleGluSerValAsnGlyLysThrValThrLeuGluLeu 642
1296 GAATTATGTTAATAGCTCCATCAAC..... 1321
643 LeuAspAsnGlyAlaGlyAlaAspAlaThrLysAspGlyValTyrSe 659
1322 TTATCTAATGCGAGTTTAACTCTCACACTTAAACGAGATGGAGTTAAAT 1371

659 rArgTyrPheThrThrTyrAspThrAsnGlyArgTyrSerValLysVala 676
1372 TAACGGTGATATTACCTCAACGAAAAATGTAATTTAACCATTAAGCAG 1421
676 tgAlaLeuGlyGlyValAlaAsnAlaAlaArgArgValIleProGlnGln 692
1422 GCTCTGGGTGATGTTTCAT.....AAAACATCACGCTGCT 1459
693 SerGlyAlaLeuTyrIleProGlyTrpIleGluAsnAspGluIleGlnTr 709
1460 ACGGGTTTTTTCATATT.....GTCGCTGGGGATCTCTAGCTTT 1500
709 pAsnProProArgProGlu..... 715
1501 TGAGAGAGAGGGCGATATAAAGCAGCTAAACGCAACAGATGCTCAAAATTACCG 1550
715IleAsnLysAspValGlnHisLys..... 724
1551 CACAAGGGACGATAACCGTCAATAAAGATGAATAACAATTTAGATTCAAT 1600
725 GlnValCysPheSerArgThrSerSerGlyGlySerPheValAlaSerAs 741
1601 AATGTATCTATTAAACGGGACGGCAAGGCTTTAAAGCTTTATTGCAATCA 1650
741 pValProAsnAlaProIleProAspLeuPheProGlyGlnIleThr. 757
1651 A.....AATAATTTCACTC 1664
759AspLeuLysAlaGluIleHisGlyGlySerLeuIleAsnLeu 771
1665 ATAAATTTGATGCGCAAAATTAACATATCTGGAATAGTAACAATTAACCAA 1714
772 Thr.....TrpThrAlaProGlyAspTyrAs 781
1715 ACCACGAAAAAGATCTTTAAATACTGGAATGCATCAAAAGACTCTTACTG 1764
781 phis.....GlyThrAlaHisLysTyr..... 788
1765 GAATGTTTCTCTTACTTTGAATACGCTGCAAAAATTTACCTTTATAA 1814
789 ..IleIleArgIleSerThrSerIleLeuAspLeuArgAspLysPheAsn 804
1815 AATTGCTTATAGCGGCTCAAAATCCCAAGATTTGAGGTCAATCAGCTAGA 1864
805 GluSerLeuGlnValAsnThrThrAlaLeuIleProLysGlu..... 818
1865 AGTTTTGCGAGCGCTACATTTTAAACGGCATCGGAGGCAAAACAACTCAA 1914
819AlaAsnSerGluGluValPheLeuPheLysProGluAsnIleT 833
1915 CATCGGAGCTAACGCAAAAGCCTTATTAAATTTAAACCAACGCGCGCTA 1964
833 hr 833
1965 CA 1966

This Page Blank (usp10)

alignment_scores: Quality: 4759.00 Length: 914
 Ratio: 5.207 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-41 x AAF81927 ..

Align seg 1/1 to: AAF81927 from: 1 to: 2745

1 MetGlyProPheLysSerValPheIleLeuIleLeuHisLeuLeuG1 17
|||||
1 ATGGGGCCATTTAAGAGTTCGTGTTTCATCTTTCACCTTCTAGA 50
|||||
17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGlu 34
|||||
51 AGGGGCCCTGAGTAATTCATCTTTCAGCTGAACAAATGGCTATGAAG 100
|||||
34 lyIleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
|||||
101 GCATTGCTGTTGCAATCGACCCCAATGTGCCAGAGATGAACACATCAIT 150
|||||
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
|||||
151 CAACAAATAAAGGACATGGTGACCCAGGCATCTCTGTATCTGTTTGAAGC 200
|||||
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGlu 84
|||||
201 TACAGGAAGCGATTTATTTTCAAAATGTTGCCATTTTGATTCCTGAAA 250
|||||
84 hrTpyLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
|||||
251 CATGGAAGACAAAGCGTGACTATGTGAGACCAAACTTGAGACCTACAAA 300
|||||
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
|||||
301 AATGCTGATGTTCTGGTGTGAGTCTACTCTCCAGGTAAATGATGAACC 350
|||||
117 oTyrThrGluGlnMetGlyAsnCysGlyGlyLysGlyGluArgIleHisL 134
|||||
351 CTACACTGAGCAGATGGCAACTGTGGAGAGAGGGTGAAGGATCCACC 400
|||||
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
|||||
401 TCACCTCTGATTTTCATTCAGGAAAAAGTTAGCTGAATATGGACCAA 450
|||||
151 GlyArgAlaPheValHisGluTyrAlaHisLeuArgTyrGlyValPheAs 167
|||||
451 GGTAGGGCATTTGCCATGAGTGGGCTCATCTACCATGGGAGTATTGGA 500
|||||
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGlnA 184
|||||
501 CGAGTACAATAATGATGAGAAATTCATCTTCAATGGAAGAAATACAAG 550
|||||
184 laValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLysCys 200
|||||
551 CAGTAAGATGTTACACGAGGTATTACTGTGTAACAAATGTAGTAAAGAGTGT 600
|||||
201 GlnGlyLysSerCysTyrThrLysArgCysThrPheAsnLysValThrG1 217
|||||
601 CAGGAGGCGAGCTGTACACCAAAAGATGCACATTCATAAAGTTACAGG 650
|||||
217 yLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrGluL 234
|||||
651 ACTCTATGAAAAGGATGTGAGTTGTTCTCCAATCCGCCAGACGGAGA 700
|||||
234 ySAlaSerIleMetPheAlaGlnHisValAspSerIleValIlePheCys 250
|||||
701 AGGCTCTATAATGTTTGCACAAACATGTTGATCTATAGTTGAATTCGT 750
|||||
251 ThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLysCy 267
|||||

751 ACAGAACAAACACACAAAGAGCTCCAAACAAAGCAAAATCAAAATG 800
|||||
267 sAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheLysL 284
|||||
801 CAATCTCGAAGACACATGGGAAGTGATCCGTGATTTCTGAGGACTTAAAG 850
|||||
284 ysThrThrProMetThrThrGlnProProAsnProThrPheSerLeuLeu 300
|||||
851 AAACCACTCTATGACACACAGCCCAATCCACCTTCTCAITGCTG 900
|||||
301 GlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySerMe 317
|||||
901 CAGATTGGACAAAGAAATGCTGTTTACTCTTGCACAAATCTGCAACAT 950
|||||
317 talatThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheL 334
|||||
951 GGGGACTGGTAACCCCTCAATCGACTGAATCAAGCAGGCCAGCTTTTCC 1000
|||||
334 euLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPheAsp 350
|||||
1001 TGCTGCAGACAGTTGAGCTGGGTGCTGGGTGGGATGGTGACATTTGAC 1050
|||||
351 SerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySerAs 367
|||||
1051 AGTGTGCCCATGTACAAAGTGAACCTACATACAGATAAACAGTGGCAGTGA 1100
|||||
367 pArgAspThrLeuAlaLysArgLeuProAlaAlaSerGlyGlyThrS 384
|||||
1101 CAGGGACACACTCGCCAAAAGATTACCTGCAGCAGCTTCAGGAGGGACGT 1150
|||||
384 erLleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLysTyr 400
|||||
1151 CCATCTGCAGCGGGCTTCATCGCGATTTACTGTGATTAGGAAGAAATAT 1200
|||||
401 ProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAsnTh 417
|||||
1201 CCAACTGATGGATCTGAATTTGCTGCTGCTGCGGATGGGGAAGACAACAC 1250
|||||
417 rIleSerGlyCysPheAsnGluValIleGlnSerGlyAlaIleIleHist 434
|||||
1251 TATAAGTGGGTGCTTTACAGAGTCAACAAAGTGGTGCATCATCCACA 1300
|||||
434 hrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluLeuSerLys 450
|||||
1301 CAGTCGCTTTGGGGCCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAA 1350
|||||
451 MetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnAsnG1 467
|||||
1351 ATGACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTTCAGAAACAATGG 1400
|||||
467 yLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValSerG 484
|||||
1401 CCTCATTTGATGCTTTTGGGGCCTTTCATCAGGAATGGAGCTGTCTCTC 1450
|||||
484 lnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSerGln 500
|||||
1451 AGCGCTCCATCCAGCTTGAGTGAAGGATTAACCTCCAGAACAGCCAG 1500
|||||
501 TrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspThrLe 517
|||||
1501 TGGATGAATGGGCACAGTGTGCTGGACAGCCCGTGGGAAGGACACTTT 1550
|||||
517 uPheLeuIleThrTrpThrThrGlnProProGlnIleLeuLeuTrpAspP 534
|||||
1551 GTTTCATTACCTGGACACAGCAGCCTCCCAAAATCCTTCTCTGGGATC 1600
|||||
534 roSerGlyGlnLysGlnGlyPheValValAspLysAsnThrLysMet 550
|||||
1601 CCAAGTGGACAGAAAGCAAGTGGGTTGTAGTGACAAAAACACCAAAATG 1650
|||||
551 AlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLysTyrse 567
|||||
1651 GCCTACCTCCAAATCCAGGCATTTGCTAAGTTGGCACTTTGGAAATACAG 1700
|||||

567 rLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSerArgAla 584
1701 TCTGCAAGCAAGCTCACAAACCTTGACCTGACTGTCTACGCTCCGCTGCGT 1750
584 erAsnAlaThrLeuProProIleThrValThrSerLysThrAsnLysAsp 600
1751 CCAATGTACCTCCCTCCCTCAATACAGTACTTCCAAAACCAAGGAC 1800
601 ThrSerLysPheProSerProLeuValValTyrAlaAsnIleArgGln 617
1801 ACCAGCAATTCGCCAGCCCTCTGCTAGTTATGCAAAATATTCGCCAAGG 1850
617 yAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerVala 634
1851 AGCTCCCAATCTCAGGGCCAGTGTACAGCCCTGATTAATCAGTGA 1900
634 snGlyLysThrValThrLeuGluLeuLeuAspAsnGlyValAlaGlyAlaAsp 650
1901 ATGGAAAACAGTTACCTTGGAACTACTGGATAATGGAGCAGGTGCTGAT 1950
651 AlaThrLysAspAspGlyValTyrSerArgTyrPheThrThrTyrAspTh 667
1951 GCTACTAAGGATGACGGTGTCTACTCAAGGTATTTTCAAACTTATGACAC 2000
667 rAsnGlyArgTyrSerValLysValAlaGAlaLeuGlyClyValAlaAsnAla 684
2001 GAATGGTATAGATACAGTGTAAAGTGCAGGCTCTGGGAGGAGTTAACGCAG 2050
684 laArgArgValIleProGlnGlnSerGlyAlaLeuTyrIleProGly 700
2051 CCAGACGAGAGTATACCCAGCAGAGTGGAGCAGCTGTACATACCTGGC 2100
701 TrpIleGluAsnAspGluIleGlnTrpAsnProProArgProGluIleAs 717
2101 TGGATTGAGAATGATGAATAACAATGGAATCCACCAAGACCTGAAATTA 2150
717 nLysAspAspValGlnHisLysGlnValCysPheSerArgThrSerSerG 734
2151 TAAGATGATGTTCAACACAGCAAGTGTGTTTCCAGCAGAACATCCTCGG 2200
734 lGlySerPheValAlaSerAspValProAsnAlaProIleProAspLeu 750
2201 GAGGCTCATTTGGCTCTGATGTCCTCCAAATGCTCCATACCTGATCTC 2250
751 PheProProGlyGlnIleThrAspLeuLysAlaGluIleHisGlyGlyse 767
2251 TTCCACCTGGCCCAATCACCGAGCTGAAGCGGAAATTCACGGGGCGAG 2300
767 rLeuIleAsnLeuThrTrpThrAlaProGlyAspAspTyrAspHisGlyT 784
2301 TCTCATTAATCTGACTTGGAGCTCTCTGGGATGATTATACCATGGAA 2350
784 hrAlaHisLysTyrIleIleArgIleSerThrSerIleLeuAspLeuArg 800
2351 CAGCTCACAATATATCATTCGAATAGTACAGTATTCCTTGATCTCAGA 2400
801 AspLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuIleProLy 817
2401 CACAAGTTCAATGAATCTCTTCAAGTGAATACTACTGCTCTCATCCCCAA 2450
817 sGluAlaAsnSerGluGluValPheLeuPheLysProGluAsnIleThrP 834
2451 GGAAGCCCAACCTCTGAGGAAGCTTTTGTAAACCCAGAAACATTAATCT 2500
834 heGluAsnGlyThrAspLeuPheIleAlaIleGlnAlaValAspLysVal 850
2501 TTGAAAATGGCAGAGATCTTTTCATTCCTATTCAGGCTGTTGATAAGGTC 2550
851 AspLeuLysSerGluIleSerAsnIleAlaArgValSerLeuPheIlePr 867
2551 GATCTGAAATCAGAAATATCCAACTTGCACGAGTATCTTTGTTTATTC 2600

867 oProGlnThrProProGluThrProSerProAspGluThrSerAlaProC 884
2601 TCCACAGACTCCGCCAGACACACCTAGCTCCTGATGAAACGCTGCTCCTT 2650
884 ysProAsnIleHisIleAsnSerThrIleProGlyIleHisIleLeuLys 900
2651 GTCCTAATATTCATATCAACAGACACCATTCCTGCGCATTCACATTTTAAA 2700
901 ileMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914
2701 ATTATGTGAAGTGGATAGGAGAACTGCAGCTGTCAATAGCC 2742

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH46124

seq_documentation_block:

ID AAH46124 standard; cDNA; 2825 BP.

XX

AC AAH46124;

DT 11-SEP-2001 (first entry)

XX

DE Human CLCA1 cDNA, SEQ ID NO:26.

XX

KW Human CLCA1; goblet cell; mouse Gob-5 orthologue; drug screening;

KW expression inhibition; antisense therapy; gene therapy;

KW chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic;

KW ss.

XX

OS Homo sapiens.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

alignment_scores:

Quality: 4759.00

Length: 914

Ratio: 5.207 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment block:
US-09-049-696-41 x AAH46124 ..
Align seg 1/1 to: AAH46124 from: 1 to: 2825

1 MetGlyProPheLysSerValPheLeuLeuLeuHisLeuLeuGlu 17
23 ATGGGGCCATTTAAGAGTCTGTGTTCATCTTGATTTCTTCACCTTCTAGA 72
17 uGlyAlaLeuSerAsnSerLeuLeuLeuGlnLeuAsnAsnGlyTyrGluG 34
73 AGGGGCCCTGAGTAATTCATCTCAGCTGAACAAATGCGTATGAAG 122
34 lyleValValAlaLeuAspProAsnValProGluAspGluThrLeuile 50
123 GCATTGCTGTCANTCGACCCCAATGTCAGAGATGAACACATCAT 172
51 GlnGlnLeuLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
173 CAACAAATAAAGGACATGTTGACCCAGGCATCTCTGTATCTGTTGAAGC 222
67 aThrGlyLysArgPheTyrPheLysAsnValAlaLeuLeuileProGluT 84
223 TACAGGAAGCGATTTTATTTTCAAAATGTTGCCATTTGATTCCTGAAA 272
84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
273 CATGGAAGACAAAGCGTACTATGTGAGACCAAACTTGAGACCTACAAA 322
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
323 AATGCTGATGTTCTGTTGCTGAGTCTACCTCCAGGTATGATGNAACC 372
117 oTyrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134
373 CTACACTGAGCAGATGGCAACTCTGGAGAGAGGGTGAAGGATCCACC 422
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
423 TCACCTCTGATTTTCATGACGAAAGTTAGCTGAATGATGGACCAAA 472
151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
473 GGTAGGGCATTTGTCATGAGTGGCTCATCTACGATGGGAGTATTTGA 522
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGlnA 184
523 CGAGTACAAATATGATGAGAAATTCATCTTATCCAAATGGAAAGATACA 572
184 laValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLysCys 200
573 CAGTAAGATGTTACACAGGTATTACTGTACAATGTAGTAAGAAGTGT 622
201 GlnGlyLysSerCysTyrThrLysArgCysThrPheAsnLysValThrGl 217
623 CAGGAGGCGAGTGTACACCAAAAGATGCACATTCATAAAGTTACAGG 672
217 yLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrGlu 234
673 ACTCTATGAAAAGGATGTAGTTGTTGTTCTCAATCCCGCCAGCGGAGA 722
234 ysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPheCys 250
723 AGGCITCTATATGTTTGACAAACATGTTGATCTATAGTTGAATTCGT 772
251 ThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLysCy 267
773 ACAGACAAACACCAACAAAGAGCTCCAAACAGCAAAATCAAAAATG 822
267 sAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheLysL 284

823 CAATCTCCGAAGCACATGGGAAGTGATCCGTGATCTGTAGGACTTTAAGA 872
284 ysThrThrProMetThrThrClnProAsnProThrPheSerLeuLeu 300
873 AAACCACTCTATGACAAACACAGCCCAATCCACCTTCTCATGTGCTG 922
301 GlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySerMe 317
923 CAGATTGGACAAACAATGTGTGTTAGTCTCTTGACAAATCTGGAAGCAT 972
317 tAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheL 334
973 GGGCAGCTGTTAAGCGCTCAATCGACTGAATCAAGCAGCGCAGCTTTTCC 1022
334 euLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPheAsp 350
1023 TGTCTGACAGCAGTGTGAGCTGGGTCTCTGGGTGGGTGGGTGACATTGAC 1072
351 SerAlaAlaHisValGlnSerGluLeuIleGlnLeuAsnSerGlySerAs 367
1073 AGTGTGCCCATGTACAAAGTGAATCATACAGATAAACAGTGGCAGTGA 1122
367 pArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyThrS 384
1123 CAGGACACACTCGCCAAAAGATTACCTGCAGCAGCTTCAGAGGAGCGT 1172
384 erIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLysTyr 400
1173 CCATCTGCAGCGGGCTTCGATCGGATTTACTGTAGTAGGAAGAAATAT 1222
401 ProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAsnTh 417
1223 CCAACTGATGATCTGAAATTTGCTGCTGACGGATGGGGAAGACAACAC 1272
417 rIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleHisL 434
1273 TATAGTGGGTGCTTTACGAGGTCAACAAAGTGGTGGCATCATCCACA 1322
434 hrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluGluLeuSerLys 450
1323 CAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAA 1372
451 MetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnAsnGl 467
1373 ATGACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCACAACAATGG 1422
467 yLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValSerG 484
1423 CCTCATTTGATGCTTTTGGGGCCCTTTTCATCAGGAAATGGAGCTGTCTCTC 1472
484 lnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSerGln 500
1473 AGCGCTCCATCCAGCTTGAGAGTAAGGATTAACCTCCAGAACAGCCAG 1522
501 TrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspThrLe 517
1523 TGGATGAATGGCACAGTGTGTCGACAGCCGCTGGGAAAGGACACTTT 1572
517 upheLeuIleThrTrpThrThrGlnProProGlnIleLeuLeuTrpAspP 534
1573 GTTTCTTATCACCTGGACACAGCGCTCCCAATCCTTCTCTGGGATTC 1622
534 roSerGlyGlnLysGlnGlyPheValValAspLysAsnThrLysMet 550
1623 CCAGTGGACAGACAGAGGTGGGCTTTGTAGTGACAAACCAACCAAAATG 1672
551 AlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLysTyrSe 567
1673 GCCTACCTCCAAATCCAGCATTTGCTAAGGTTGGCACTTGGAAATACAG 1722
567 rLeuGlnAlaSerSerClnThrLeuThrLeuThrValThrSerArgAlaS 584

1723 TCTGCAAGCAAGCTCACAAACCTTGACCTGACTGTCACGTCCTCCCTGCGT 1772
584 eRAsnAlaThrLeuProProlleThrValThrSerLysThrAsnLysAsp 600
1773 CCAATGCTACCGCTGCTCCTCAATATACAGTGACTTCCAAAACGAACAGAC 1822
601 ThrSerLysPheProSerProLeuValValThrAlaAsnIleArgGlnG1 617
1823 ACCACAAATTCCTCCAGCCCTCTGGTAGTTATGCAAAATATTCGCCAAGG 1872
617 yAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerVala 634
1873 AGCTCCCCCAATCTCAGGCCAGGTGTACAGCCCTGATTGAATCAGTGA 1922
634 sNGLyLysThrValThrLeuGluLeuLeuAspAsnGLyAlaGLyAlaAsp 650
1923 ATGGAACCAACAGTTACCTTGGAACTACTTGGATAATGGAGCAGGTGCTGAT 1972
651 AlaThrLysAspAspGLyValTyrSerArgTyrPheThrThrTyrAspTh 667
1973 GCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCAAACTTATGACAC 2022
667 rAsnGLyArgTyrSerValLysValArgAlaLeuGLyGLyValAsnAlaA 684
2023 GAATGTAGATACAGTGTAAAGTCCGGCTCTGGAGGAGTTACGCAG 2072
684 laArgArgValIleProGlnGlnSerGLyAlaLeuTyrIleProGLy 700
2073 CCAGACGAGAGTGTATCCCCAGCAGAGCTGGAGCAGCTGTACATACCTGGC 2122
701 TrpIleGluAsnAspGluIleGlnTrpAsnProProArgProGluIleAs 717
2123 TGGATTTGAGAATGATGAATACAATGGAATCCACCAAGACCTGGAATTA 2172
717 nLysAspValGlnHisLysGlnValCysPheSerArgThrSerSerG 734
2173 TAAGATGATGTTCAACACAGCAAGTGTGTTTCAGCAGACATCTCGG 2222
734 lyGLySerPheValAlaSerAspValProAsnAlaProIleProAspLeu 750
2223 GAGGCTCATTTGTGGCTTCTGATGTCCCAATATCCATCCATACCTGATCTC 2272
751 PheProGLyGlnIleThrAspLeuLysAlaGluIleHisGLyGLySe 767
2273 TTCCACCTGGCCAAATCAGCAGCTGAAGCGGAAATTCACGGGGGCGAG 2322
767 rLeuIleAsnLeuThrTrpThrAlaProGLyAspAspTyrAspHisGLyT 784
2323 TCTCATTAATCTGACTTGGACAGCTCCTGGGATGATGATGATGACCATGGA 2372
784 hAlaHisLysTyrIleIleArgIleSerThrSerIleLeuAspLeuArg 800
2373 CAGCTCACAAAGTATATATATTCGAATTAAGTACAAAGTATTTCTTGTATCTCAGA 2422
801 AspLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuIleProLy 817
2423 GACAAGTTCAATGAATCTCTCAAGTGAATACTACTGCTCTCATCCCCAA 2472
817 sGluAlaAsnSerGluGluValPheLeuPheLysProGluAsnIleThrP 834
2473 GGAAGCCCAACTCTGAGGAAGTCTTTTCTTTAAACCCAGACAAACATTA 2522
834 heGluAsnGLyThrAspLeuPheIleAlaIleGlnAlaValAspLysVal 850
2523 TTGAAATGGCAGACAGATCTTTTTCATTGCTATTCAGGCTGTTGATAAGGTC 2572
851 AspLeuLysSerGluIleSerAsnIleAlaArgValSerLeuPheIlePr 867
2573 GATCTGAATCAGAAATATCCAAATTCACAGAGTATCTTTGTTTATTTCC 2622
867 oProGlnThrProProGluThrProSerProAspGluThrSerAlaProC 884
2623 TCCACAGACTCCGCCAGACAGACCTAGTCTCTGATGAACAGTCTGCTCCTT 2672

884 ysProAsnIleHisIleAsnSerThrIleProGLyIleHisIleLeuLys 900
2673 GTCTTAATATTATCATATCAACGACCATTCCTGTCATTTCACATTTTAAAA 2722

901 IleMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914
2723 ATTATGTGAAGTAGTAGGAGAACTGAGCTGTCAATAGCC 2764

seq_name: /SDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH34879

seq_documentation_block:

ID: AAH34879 standard; cDNA; 2854 BP.

XX AAH34879;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen encoding cDNA SEQ ID NO:1961.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

KW colorectal carcinoma; chromosome 1; ss.

XX Homo sapiens.

XX WO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

DR WPI; 2001-235357/24.

DR P-PSDB; AAG75474.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers -

PS Claim 1; Page 3462-3463; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patients own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAB7789 represent sequences used in the exemplification of the present invention.

CC N.H. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for

CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX SQ Sequence 2854 BP; 885 A; 628 C; 643 G; 698 T; 0 other;

alignment_scores:

Quality: 4759.00

Ratio: 5.207

Percent Similarity: 100.000

Percent Identity: 100.000

Length: 914

Gaps: 0

alignment_block:

US-09-049-696-41 x AAH34879 ..

Align seg 1/1 to: AAH34879 from: 1 to: 2854

1 MetGlyProPheLysSerValPheIleLeuIleLeuHisLeuLeuG1 17
|||||
35 ATGGGCCATTAAAGAGTCTGTGTTCAATCTTGATTTCACTTCTAG 84
17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34
|||||
85 AGGGGCCCTGAGTAATCACTATTCACTGAGCTGAACAATAATGGCTATGAAG 134
34 lyIleValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
|||||
135 GCATTTGCTGTGAATGCAATGAGCCCAATGTGCCAGAAAGATGAAACACTCAATT 184
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
|||||
185 CAACAAATAAAGGACATGGTGCACCCAGGCATCTCTGTATCTGTTTGAAGC 234
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluT 84
|||||
235 TACAGGAAGCGATTTTATTTCAAAATGTTGCCATTTTGATTCCTGAAA 284
84 hTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
|||||
285 CATGGAAGACAAAGGCTGACTATGTGAGACCAAACTTGAGACCTACAAA 334
101 AsnAlaAspValLeuValAlaGluSerThrProProGlyAsnAspGluPr 117
|||||
335 AATGCTGATGTTCTGTTGCTGAGTCTACTCTCCAGGTAATGATGAACC 384
117 oTyrThrGluGlnMetGlyAsnGlyGluLysGlyGluArgIleHisL 134
|||||
385 CTACACTGAGCAGATGGCACTGTGGAGAGAGGGTGAAGGATCCACC 434
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
|||||
435 TCACCTCTGATTTTCATTTGCAGGAAAAAAGTTAGCTGAATATGAGACCACAA 484
151 GlyArgAlaPheValHisGluTyrAlaHisLeuArgTrpGlyValPheAs 167
|||||
485 GCTAGGCGCATTTGCCATGAGTGGGCTCATCTACGATGGGAGTATTGTA 534
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGlnA 184
|||||
535 CGAGTACATAATGATGAGAAATCTACTTATCCAATGGGAATACAAAG 584
184 laValArgCysSerAlaGlyIleThrGlyThrAsnValValLysCys 200
|||||
585 CAGTAAGATGTTACGACAGGTATTACTGGTCAAAATGAGTAAAGAAGTGT 634
201 GlnGlyGlySerCysTyrThrLysArgCysThrPheAsnLysValThrG1 217
|||||
635 CAGGAGGCGACTGTTACCCAAAAGATGACATTCAATAAAGTAACAGG 684
217 yLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrGluL 234
|||||
685 ACTCTATGAAAAGGATGTGAGTTTGTCTCCAATCCGCCAGACGGAGA 734
234 ysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPheCys 250
|||||
735 AGGCTTCTATAATGTTGCAACATGTTGATTCTATAGTTGAATTTCTGT 784
251 ThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLysCy 267
|||||
785 ACAGAACAAAAACCACAAAGAGAGCTCCAAACAAAGCAAAATCAAAAATG 834
267 sAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheLysL 284
|||||
835 CAATCTCCGAAGCACATGGGAAGTGATCCGTGATTCTGAGGACTTTAAGA 884

284 ysThrThrProMetThrThrGlnProProAsnProThrPheSerLeuLeu 300
|||||
885 AAACCACTCTTATGACAACACAGCCACCAAAATCCACACTTCTCAATTGCTG 934
301 GlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySerMe 317
|||||
935 CAGATTGGACAAAGAAATTTGTGTTTAGTCTTGACAAATCTGGAAGCAT 984
317 tAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheL 334
|||||
985 GCGGACTGGTAACCGCTCAATCGACTGAATCAAGCAGCCAGCTTTTCC 1034
334 euLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPheAsp 350
|||||
1035 TGCTCGACACAGTTTCAGCTGGGCTCTGGGTTGGGATGTTGACATTTGAC 1084
351 SerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySerAs 367
|||||
1085 AGTGCTGCCCATGTACAAAGTGAATCATACAGATAAACAGTGGCAGTGA 1134
367 pArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyThrS 384
|||||
1135 CAGGGACACACTCGCCAAAAGATTACCTGCAGCAGCTTCAGGAGGAGCT 1184
384 erIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLysTyr 400
|||||
1185 CCATCTGCAGCGGCTTCGATCGCATTTACTGTGATTAGGAAGAAATAT 1234
401 ProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAsnTh 417
|||||
1235 CCAACTGATGGATCTGAATTTGTGCTGTCAGGATGGGAAGACACAC 1284
417 rIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleHisT 434
|||||
1285 TATAAGTGGTGTCTTTACAGAGTCAAAAGTGGTGCATCATCCACA 1334
434 hrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluLeuSerLys 450
|||||
1335 CAGTCTGCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAA 1384
451 MetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnAsnG1 467
|||||
1385 ATGACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTTCAGAACAAATGG 1434
467 yLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValSerG 484
|||||
1435 CCTCATTTGATGCTTTTGGGGCCCTTTTCATCAGGAAATGGAGCTGTCTCTC 1484
484 lnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSerGln 500
|||||
1485 AGCGCTCCATCCAGCTTGAGAGTAAGGGATTAACCTCCAGAACACCCAG 1534
501 TrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspThrLe 517
|||||
1535 TGGATGAATGGCACAGTGTCTGGACACGACCGTGGGAAAGGACACTTT 1584
517 upheLeuIleThrTrpThrThrGlnProProGlnIleLeuLeuThrTrpAspP 534
|||||
1585 GTTTCCTTATCACCTGGACACGAGGCTCCCCAAATCCTTCTCTGGGATC 1634
534 roSerGlyGlnLysGlnGlyPheValValAspLysAsnThrLysMet 550
|||||
1635 CCAGTGGACAGAAAGAGGTGGCTTTGTAGTGGACAAAACACCAAAATG 1684
551 AlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLysTyrSe 567
|||||
1685 GCCTACTCTCCAAATCCAGCATTTGTAAGTGGCAGCTTGGAAATACAG 1734
567 rLeuGlnAlaSerSerGlnThrLeuThrValThrSerArgAla 584
|||||
1735 TCTGCAAGCAAGCTTCAAAACCTTGACCTTGACCTGTACGCTCCCGTCCGCT 1784

584 erAsnAlaThrLeuProIleThrValThrSerLysThrAsnLysasp 600
1785 CCAATGCTACCTGCTGCTCAATACAGTACTCCAAAACGAACAGGAC 1834
601 ThrSerLysPheProSerProLeuValValThrAlaAsnIleArgGlnG1 617
1835 ACCACCAATTCCTCCAGCCCTCTGGTAGTTATGCAATATTCGCAAGG 1884
617 yAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerVala 634
1885 AGCTCCCTCCCAATTCACAGGCGCAGTGTACAGCCCTGATTAATCAGTGA 1934
634 sNGLyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAlaasp 650
1935 ATGGAAAACAGTTACCTTGAACACTACGAGTATGAGAGAGGTGCTGAT 1984
651 AlaThrLysAspAspGlyValTyrSerArgTyrPheThrThrTyrAspTh 667
1985 GCTACTAAGGATGACGGTGTCTACTCAAGGTATTTACAACTTATGACAC 2034
667 rAsnGlyArgTyrSerValLysValArgAlaLeuGlyGlyValAsnAlaA 684
2035 GAATGGTAGATACAGTGTAAAGTGGCGCTCTGGAGAGGTTACGCGAG 2084
684 laAtgArgArgValIleProGlnGlnSerGlyAlaLeuTyrIleProGly 700
2085 CCAGACGGAGAGTGTATCCAGCAGAGAGTGGAGCACTGTACATACCTGCG 2134
701 TrpIleGluAsnAspGluIleGlnTrpAsnProProArgProGluIleAs 717
2135 TGGATTGAGATGATGAATAATCAATGAATCCACCAAGACCTGAAATTA 2184
717 nLysAspValGlnHisLysGlnValLysPheSerArgThrSerG 734
2185 TAAGATGATGTTCACCAAGCAAGTGTGTTCAGCAGAACATCCTCGG 2234
734 lyGlySerPheValAlaSerAspValProAsnAlaProIleProAspLeu 750
2235 GAGGCTCATTTGTGGCTCTGATGTCCTCCCAATGCTCCATACCTGATCTC 2284
751 PheProGlyGlnIleThrAspLeuLysAlaGluIleHisGlyGlyse 767
2285 TTCCACCTGCGCAATCAGCAGCTGAAGCGGAATTCACGGGGGCGAG 2334
767 rLeuIleAsnLeuThrTrpThrAlaProGlyAspAspTyrAspHisGlyT 784
2335 TCTCATTAATCTGACTGGAGCTCCTGGGGATGATTATGACCATGAA 2384
784 hrAlaHisLysTyrIleIleArgIleSerThrSerIleLeuAspLeuArg 800
2385 CAGCTCACAGATATATCATTCGAATAGTACAGTATCTTGATCTCAGA 2434
801 AspLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuIleProLy 817
2435 GACAAGTTCAATGAATCTCTTCAAGTGAATACTACTGCTCTCATCCCAA 2484
817 sGluAlaAsnSerGluGluValPheLeuPheLysProGluAsnIleThrP 834
2485 GGAAGCCCAACTCTGAGGAAGTCTTTTGTAAACCCAGAAAACATTA 2534
834 heGluAsnGlyThrAspLeuPheIleAlaIleGlnAlaValAspLysVal 850
2535 TTGAAAATGGCAGACATCTTTTCATGCTATTCAGGCTGTTGATNAGTCT 2584
851 AspLeuLysSerGluIleSerAsnIleAlaArgValSerLeuPheIlePr 867
2585 GATCTGAAATCAGAAATATCAACATTCACAGAGTATCTTTGTTTATTC 2634
867 oProGlnThrProProGluThrProSerProAspGluThrSerAlaProC 884
2635 TCCACAGACTCCGCCAGAGACACCTAGTCTCGATGAACAGCTGCTGCTCT 2684
884 ysProAsnIleHisIleAsnSerThrIleProGlyIleHisIleLeuLys 900

2685 GTCTCTAATATTCATATCAACAGCACCATCTCTGCGATTCACATTTAAAA 2734
901 IleMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914
2735 ATTATGTCGAAGTGGATAGGAGAACTGCGCTGTCATAGCC 2776
seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF81787
seq_documentation_block:
ID AAF81787 standard; cDNA; 2854 BP.
XX AAF81787;
AC AAF81787;
DT 12-JUN-2001 (first entry)
XX Human secreted protein gene 1 SEQ ID NO:11.
DE Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW immunostimulant; cytostatic; cardiac; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
KW antialzheimer; antiparkinsonian; antimicrobial; vulnary; gene therapy;
KW immune disorder; hyperproliferative disorder; cardiovascular disease;
KW cancer; angiogenic disorder; neurological disorder; infectious disease;
KW wound healing; regeneration; chemotaxis; chromosome 1; ss.
XX Homo sapiens.
OS
XX
XX WO200112775-A2.
PN
XX
XX 22-FEB-2001.
PD
XX
XX 16-AUG-2000; 2000WO-US22325.
PF
XX
XX 17-AUG-1999; 9905-0149182.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ni J, Florence KA, Fiscella M, Wei P, Baker KP;
PI Elrse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;
XX WPI; 2001-147550/15.
XX P-PSDB; AAB74733.
DR
XX
XX Nucleic acids encoding 25 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
PT
XX
XX Claim 1; Page 441; 485pp; English.
PS
XX
XX AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733
CC to AAB74772. Human secreted proteins can have activities based on the
CC tissues and cells they are expressed in. Example of activities include:
CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;
CC antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiac;
CC vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic;
CC anticonvulsant; antialzheimer; antiparkinsonian; antimicrobial; and
CC vulnary. Human secreted proteins can be used in gene therapy and
CC vaccine. Human secreted protein nucleotide sequences (NAM1) and proteins
CC (PEP1) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. For example, NAM1
CC and PEP1 may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patients genome
CC that affect the activity of proteins by expressing inactive proteins or
CC to supplement the patients own production of polypeptides. Disorders that
CC may be prevented, diagnosed and/or treated include immune disorders,
CC hyperproliferative disorders (e.g. cancers), cardiovascular diseases,
CC angiogenic disorders, neurological disorders, infectious diseases and/or
CC for promoting wound healing, regeneration and /or chemotaxis. AAF81778 to
CC AAF81786 and AAB74732 represent sequences used in the exemplification of
CC the present invention.
XX

SQ Sequence 2854 BP; 885 A; 628 C; 643 G; 698 T; 0 other;

alignment_scores:
Quality: 4759.00 Length: 914
Ratio: 5.207 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-41 x AAF81787 ..

Align seg 1/1 to: AAF81787 from: 1 to: 2854

1 MetClyProPheLysSerValPheLeuLeuLeuHisLeuLeuG1 17
35 ATGGGGCCATTAAAGATTCTGTGTCATCTTCACTTTCACCTTCTAGA 84
17 uGlyAlaLeuSerAsnSerLeuLeuGlnLeuAsnAsnGlyTyrGluG 34
85 AGGGCCCTGAGTAATTCATCTATTAGCTGAACAACAAATGGCTATGAAG 134
34 lytleValValAlaIleAspProAsnValProGluAspGluThrLeuLe 50
135 GCATTGTGCTGCAATCGACCCCAATGTGCCAAGATGAACACACTCAT 184
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluA 67
185 CAACAATAAAGGACATGGTGCCAGCCAGCATCTGTGTATCTGTTGAAGC 234
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuLeuProGlu 84
235 TACAGGAAGCGATTTATTTCAAAAATGTTGCCATTTTGTATCTCTGAAA 284
84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
285 CATGGAAGACAAAGCGCTGACTATGTGAGACCAAACTTGAGACCTACAA 334
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
335 AATGCTGATGTTCTGTTGCTGAGTCTACTCTCCAGGTAATGATGAACC 384
117 oTyrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134
385 CTACACTGAGCAGATGGGCAACTGTGGAGAGAGGGTGAAAGATCCACC 434
134 euThrProAspPheIleAlaGlyLysLeuAlaGluTyrGlyProGln 150
435 TCACCTCTGATTTTCATTTCAGGAAAAAAGTTAGCTGAATATGGACCACAA 484
151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
485 GGATAGGCAATTTGTCATGAGTGGGCTCATCTACGATGGGAGATTTGA 534
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGlnA 184
535 CGAGTACATATATGATGAGAAATTTCTACTTATCCAAATGGAAGATACAAG 584
184 laValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLysCys 200
585 CAGTAAGATGTTTCAGCAGGATATGCTGTACAAATGTATGAAGAAGTGT 634
201 GlnGlyGlySerCysTyrThrLysArgCysThrPheAsnLysValThrG 217
635 CAGGGAGGCGAGCTGTACACCAAAAGATGACATTCATTAAGTAACAGG 684
217 yLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrGluL 234
685 ACTCTATGAAAAGGATGTGAGTTTGTCTCCAATCCCGCCAGACGGAGA 734
234 ysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPheCys 250
735 AGGCTTCTATAATGTTTGGACACACATGTTGATTTCTATAGTTGAATTC 784

251 ThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLysCy 267
785 ACAGACAAACACCAACAAAGAGCTCCAAACAAAGCAAAATCAAAATG 834
267 sAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheLysL 284
835 CAATCTCCGAAGACATCGGNAAGTATCCGCTGATTCTGAGGACTTTAAGA 884
284 ysThrThrProMetThrThrGlnProAsnProThrPheSerLeuLeu 300
885 AAACCACTCTATGACAACACAGCCACCAAAATCCACCTTCTCATTTGCTG 934
301 GlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySerMe 317
935 CAGATTGGACAAAGAAATTTGTGTTTGTAGTCTTGACAAATCTGGAAGCAT 984
317 tAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheL 334
985 GCGACTGTGTAACCGCTCATCGACTGAATCAAGCAGCGCCAGCTTTTCC 1034
334 euLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPheasp 350
1035 TGCTGCAGACAGATTGAGCTGGGCTCTGGGTGGGATGGTGACATTTGAC 1084
351 SerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySerAs 367
1085 AGTGTCGCCATGTACAAAGTGAACCTATACAGATAACACATGCGCAGTGA 1134
367 pArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyThrS 384
1135 CAGGACACACTCGCCAAAAGATTACCTGCAGCAGCTTCAGGAGGGAGCT 1184
384 erIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLysTyr 400
1185 CCATCTGCAGCGGCTTCGATCGGCATTTACTGTGATTAGGAAGAAATAT 1234
401 ProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAsnTh 417
1235 CCAACTGATGGATCTCAAAATTTGCTGCTGCGGATGGGAAGACAACAC 1284
417 rIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleHisT 434
1285 TATAAGTGGTGTCTTTACGAGGTCAAAAGTGGTGGCCATCATCCACCA 1334
434 hrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluLeuSerLys 450
1335 CAGTCCTTTGGGCGCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAA 1384
451 MetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnGln 467
1385 ATGACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTTCAGAAACAATGG 1434
467 yLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValSerG 484
1435 CCTCATTTGATGCTTTTGGGCGCTTTCATCAGGAATGAGCTGTCTCTC 1484
484 lnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSerGln 500
1485 AGCGCTCCATCCAGCTTGAGATGAGGATTAACCTCCAGAACAGCCAG 1534
501 TrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspThrLe 517
1535 TGGATGAATGGCAGTGATCGTGGACAGCCCGCTGCCAAATCCTTCTCTGGGATC 1584
517 uPheLeuIleThrTrpThrThrGlnProProGlnIleLeuLeuTrpAspp 534
1585 GTTCTTATACCTGGACAACGAGCGCTCCCAAAATCCTTCTCTGGGATC 1634
534 roSerGlyGlnLysGlnGlyPheValValAspLysAsnThrLysMet 550
1635 CCAGTGGACAGAAAGAGGCTGCTTTGTAGTGGACAAAAACCAAAATG 1684
551 AlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLysTyrse 567

```
|||||
1685 GCCTACCTCCAAATCCCAAGGATTGCTAAGGTGGCACTTGGAAATACAG 1734
567 rLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSerArgAla 584
|||||
1735 TCTGCAAGCAAGCTCACAAACCTTGACCTGACTGCTACGCTCCGCTGGT 1784
584 eAsnAlaThrLeuProProIleThrValThrSerLysThrAsnLysAsp 600
|||||
1785 CCAATGCTACCTCGCTCCAAATTACAGTGACTTCCAAAACGAACAGGAC 1834
601 ThrSerLysPheProSerProLeuValValTyrAlaAsnIleArgGlnG 617
|||||
1835 ACCAGCAAAATCCCAAGCCCTCTGGTAGTTTATGCAAAATATTCGCCAAG 1884
617 yAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerVal 634
|||||
1885 AGCCTCCCAATTTCTCAGGGCCAGTGTCACAGCCCTGATTGAATCAGTGA 1934
634 sNGLyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAlaAsp 650
|||||
1935 ATGGAAAACAGTTACCTTGGAACTACTGGATAATGGAGCAGGTGCTGAT 1984
651 AlaThrLysAspAspGlyValTyrSerArgTyrPheThrThrTyrAspTh 667
|||||
1985 GCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCACTCACTTATGACAC 2034
667 rAsnGlyArgTyrSerValLysValArgAlaLeuGlyGlyValAlaAsnAla 684
|||||
2035 GAATGGTAGATACAGTGTAAAAGTCGGGCTCTGGGAGGAGTTAACGCAG 2084
684 laArgArgArgValIleProGlnGlnSerGlyAlaLeuTyrIleProGly 700
|||||
2085 CCAGACGGAGAGTGTATACCCAGCAGAGCTGGAGCACTGTACATACCTGCG 2134
701 TrpIleGluAsnAspGluIleGlnTrpAsnProProArgProGluIleAs 717
|||||
2135 TGGATTGAGATGATGAATACATGATGAATCCACCACACCTGGAATTA 2184
717 nLysAspAspValGlnHisLysGlnValIcysPheSerArgThrSerSerG 734
|||||
2185 TAAGATGATGTTCACCAAGCAAGTGTGTTTCAGCAGACATCCTCGG 2234
734 lyGlySerPheValAlaSerAspValProAsnAlaProIleProAspLeu 750
|||||
2235 GAGGCTCATTTGTGGCTTCTGATGTCCTCCAAATGCTCCCATCCTGATCTC 2284
751 PheProGlyGlnIleThrAspLeuLysAlaGluIleHisGlyGlySer 767
|||||
2285 TTCCACCTGGCCAAATCACCAGCTGAAGGGGAATTCACGGGGGCGAG 2334
767 rLeuIleAsnLeuThrTrpThrAlaProGlyAspAspTyrAspHisGlyT 784
|||||
2335 TCTCTAATATCTGACTTGGACAGCTCTCTGGGATGATTATGACCATGGA 2384
784 hAlaHisLysTyrIleIleArgIleSerThrSerIleLeuAspLeuArg 800
|||||
2385 CAGCTCACAAAGTATATCATCTCGAATPAAGTACAAGTATTCTTCATCTCAGA 2434
801 AspLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuIleProLy 817
|||||
2435 GACAAGTTCAATGAATCTCTTCNAGTGAATCTACTGCTCTCATNCCCAA 2484
817 sGluAlaAsnSerGluGluValPheLeuPheLysProGluAsnIleThrP 834
|||||
2485 GGAAGCCCAACTCTGAGGAAGCTTTTCTTTAAACCCAGAAAACATTACTT 2534
834 hGluAsnGlyThrAspLeuPheIleAlaIleGlnAlaValAspLysVal 850
|||||
2535 TTGAAAATGGCAGACATCTTTTCATTGCTATTTCAGGCTGTTTGATAAGGTC 2584
851 AspLeuLysSerGluIleSerAsnIleAlaArgValSerLeuPheIlePr 867
|||||
```

```
2585 GATCTGAATCAGAAATATCCAAACATTCACAGATATCTTTGTTTATTC 2634
867 oProGlnThrProGluThrProSerProAspGluThrSerAlaProC 884
|||||
2635 TCCACAGACTCCGCCAGAGACACACCTAGTCTGTGATGAAACGCTGCTCCTT 2684
884 yProAsnIleHisIleAsnSerThrIleProGlyIleHisIleLeuLys 900
|||||
2685 GTCTTAATATTATATCAATCAACGACCACTTCTGTCATTCACATTTAA 2734
901 IleMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914
|||||
2735 ATTATGTGAAGTGGATAGGAGAACTGCGCTGTCATAGCC 2776
seq_name: /SDS2/gcdata/geneseq/geneseq/NA1999.DAT:AAZ09840
seq_documentation_block:
ID_AAZ09840 standard; cDNA; 3111 BP.
XX_AAZ09840;
XX
XX 26-NOV-1999 (first entry)
XX
DE Human membrane spanning protein MSP-5 cDNA fragment 2.
XX
XX Membrane spanning protein; MSP; human; treatment; diagnosis; prevention;
KW neoplastic disorder; immunological disorder; reproductive disorder;
KW MSP-5; ds.
XX
XX Homo sapiens.
XX
XX WO9946380-A2.
XX
XX 16-SEP-1999.
XX
XX 09-MAR-1999; 99WO-US05073.
XX
XX 13-MAR-1998; 98US-0039064.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Taug YT, Bandman O, Lal P, Hillman JL, Yue H, Corley NC;
PI Guegler KJ, Kaser MR, Baughn MR, Shah P;
XX
XX WFi; 1999-551409/46.
DR P-PSDB; AAY33298.
XX
XX New human membrane spanning proteins used to, e.g. prevent and treat
neoplastic disorders.
XX
XX Example 1; Page 80-81; 81pp; English.
XX
XX This invention describes novel human membrane spanning proteins (MSPs),
and the polynucleotides encoding them. The products of the invention are
used to diagnose, prevent and treat neoplastic, immunological and
reproductive disorders. This sequence encodes a human membrane spanning
protein MSP-5 fragment.
XX
XX Sequence 3111 BP; 968 A; 664 C; 691 G; 788 T; 0 other;

alignment_scores:
Quality: 4759.00 Length: 914
Ratio: 5.207 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-649-696-41 x AAZ09840
Align seg 1/1 to: AAZ09840 from: 1 to: 3111

1 MetGlyProPheLysSerValPheIleLeuIleLeuHisLeuLeuG1 17
|||||
```

34 ATGGGCCATTTAAGAGTTCTGTGTTTCATCTTGATTCTTCACCTTCTAGA 83
17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34
84 AGGGGCCCTGAGTAATTCATCTTACGCTGAACAAATGGCTATGAAG 133
34 lytleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
134 GCATTGTGCTGCAATCGACCCCAATGTGCCAGAGATCAACACTCAT 183
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
184 CAACAATAAGGACATGTTGACCCAGGCATCTCTGATCTGTTTGAAGC 233
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluT 84
234 TACAGGAACGCAATTTATTTCAAAATGTTGCCATTTTGATTCCCTGANA 283
84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
284 CATGGAAGACAAAGCTGACTATGTGAGACCAAACTTGAGACCTACAAA 333
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
334 AATGCTCATCTCTGCTGTGCTGAGTCTACTCTCCAGGTAATCATGAACC 383
117 oTyrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134
384 CTACACTGACGAGTGGCAACTGTGGAGAGAAGGGTGAAGGATCCACC 433
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
434 TCACCTCTGATTTTCATTCGAGGAAAAAGTTAGCTGAATATGGACCACAA 483
151 GlyArgAlaPheValHisGluTyrAlaHisLeuArgTrpGlyValPheAs 167
484 GGTAGGCGATTGTCCCATGAGTGGGCTCATCTACGATGGGGAGTATTGA 533
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGlnA 184
534 CGAGTACAAATATCATGAGAAATCTACTTATCCAATGGAAGAATACAAAG 583
184 laValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLysCys 200
584 CAGTAAGATGTTACGACAGGATTTACTGGTACAAATGTAGTAAAGAGTGT 633
201 GlnGlyGlySerCysTyrThrLysArgCysThrPheAsnLysValThrGl 217
634 CAGGAGGCACTGTTACACCAAAAGATGCACATTCATTAAGTAACAGG 683
217 yLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrGluL 234
684 ACTCTATGAAAAGGATGTGAGTTGTTCTCCAATCCCGCCAGACGGAGA 733
234 ysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPheCys 250
734 AGGCTTCATATATGTTGCACACACATGTTGATCTATAGTTGAATCTGT 783
251 ThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLysCy 267
784 ACAGAAACAAACCAACAAAGAGCTCCAAACAAAGCAAAATCAAAAATG 833
267 sAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheLysL 284
834 CAATCTCGAAGACATGGGAAGTATCCGTTGATTCGAGGACATTAAGA 883
284 ysThrProMetThrThrGlnProProAsnProThrPheSerLeuLeu 300
884 AAACCACTCTATGACACACAGCACCACCAATCCACCTTCTCATTTGCTG 933
301 GlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySerMe 317
934 CAGATTGGACAAAGAATTTGTGTTTAGTCTCTGACAAATCTGGAAGCAT 983

317 talThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheL 334
984 GGCACCTGGTAACCGCCTCAATCGACTGAATCAAGCAGCGCAGCTTTTC 1033
334 euLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPheAsp 350
1034 TGCCTCGACACAGTTTCAGCTGGGGTCTCGGTTGGGATCGTGACATTTGAC 1083
351 SerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySerAs 367
1084 AGTGTGCCCATGTACAAAGTGAACCTCATACGATAAACAAGTGGCAGTGA 1133
367 pArgAspThrLeuAlaLysArgLeuProAlaAlaSerGlyGlyThrS 384
1134 CAGGACACACTCGCCAAAAGATTACCTGCAGCAGCTTCAGGAGGACGT 1183
384 erIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLysTyr 400
1184 CCATCTGACGGGGCTTCGATCGGCATTTACTGTGATTAGGAAGAAATAT 1233
401 ProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAsnTh 417
1234 CCAACTGATGGATCTGAAATTTGTGCTGCTGACGGATGGGAAGACAAC 1283
417 rIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleHist 434
1284 TATAAGTGGGTGCTTTAAGCAGGTCAAAAGTGTGCCATCATCCACA 1333
434 hrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluLeuSerLys 450
1334 CAGTCCGCTTTGGGGCCCTCTCGAGCTCAAGACTAGACGAGCTGTCCAAA 1383
451 MetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnGln 467
1384 ATGACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTTCAGAACATGG 1433
467 yLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValSerG 484
1434 CCTCATGATGCTTTTGGGGCCCTTTCATCAGGAATGAGCTGTCTCTC 1483
484 InArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSerGln 500
1484 AGCGCTCCATCCAGCTTGAGAGTAAAGGATTAACCCCTCCAGAACAGCCAG 1533
501 TrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspThrLe 517
1534 TGGATGAATGGCACAGTGTATCGTGACACGCCGTGGGAAGGACACTTT 1583
517 uPheLeuIleThrTrpThrThrGlnProGlnIleLeuLeuLeuTrpAspP 534
1584 GTTCTTATCACCTGGACACAGCAGCCTCCCAAAATCCTTCTCTGGGATC 1633
534 roSerGlyGlnLysGlnGlyGlyPheValValAspLysAsnThrLysMet 550
1634 CCAGTGGACAGCAAGGTTGGTGTAGTGGACACGCCGTGGGAAGGACACT 1683
551 AlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLysTyrSe 567
1684 GCCTACTCCAAATCCAGGCATGCTAGGTTGGCACTTGGNAATACAG 1733
567 rLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSerArgAlaS 584
1734 TCTGCAAGCAAGCTCACAAACCTTGACCTCCCAAAATCCTTCTCCGTCGT 1783
584 erAsnAlaThrLeuProProIleThrValThrSerLysThrAsnLysAsp 600
1784 CCAATGCTACCTCCCTCCAAATTTACAGTGTCTCCAAAGCAAGGAC 1833
601 ThrSerLysPheProSerProLeuValValTyrAlaAsnIleArgGlnI 617
1834 ACCAGCAATTCGCCAGCCCTCTGGTAGTTTATGCAAAATATTCCCAAGG 1883


```
617 yAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerVala 634
|||||
1884 AGCTCCCAATCTCAGGCCAGTGTACAGCCCTGATTGAATCAGTGA 1933
|||||
634 snGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAlaAsp 650
|||||
1934 ATGGAACAACAGTTACCTTGGAACTACTGGATATGAGCAGGTGCTGAT 1983
|||||
651 AlaThrLysAspAspGlyValThrSerArgThrPheThrThrAspTh 667
|||||
1984 GCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCAACAATTTATGACAC 2033
|||||
667 rAsnGlyArgThrSerValLysValArgAlaLeuGlyGlyValAsnAlaA 684
|||||
2034 GAATGGTAGATACACTGTAAAGTCCGGCTCTGGAGAGGTAAACCGAG 2083
|||||
684 laArgArgValIleProGlnGlnSerGlyAlaLeuTyIleProGly 700
|||||
2084 CCAGACGAGAGTGATACCCCGCAGCAGAGTGGAGCACTGTACATACCTGGC 2133
|||||
701 TrpIleGluAsnAspGluIleGlnTrpAsnProProArgProGluIleAs 717
|||||
2134 TGGATTGAGATGATGAATACAAATGGAATCCACCAAGACCTGAAATTA 2183
|||||
717 nLysAspAspValGlnHisLysGlnValCysPheSerArgThrSerSerg 734
|||||
2184 TAAGCATGATGTTCAACACAAAGCAAGTGTGTTCCAGCAGAACATCTCCGG 2233
|||||
734 lyGlySerPheValAlaSerAspValProAsnAlaProIleProAspLeu 750
|||||
2234 GAGGCTCATTTGTGGCTTCTGATGTCCCAATGCTCCCATACCTGATCTC 2283
|||||
751 PheProGlyGlnIleThrAspLeuLysAlaGluIleHisGlyGlySe 767
|||||
2284 TTCCACCTCGCCAAATCACCGACCTGAAGCGGAAATTCACGGGGCGAG 2333
|||||
767 rLeuIleAsnLeuThrTrpThrAlaProGlyAspAspTyAspHisGlyT 784
|||||
2334 TCTCATTAATCTGACTGGACAGCTCTCTGGGATGATTATGACCATGGAA 2383
|||||
784 hrAlaHisLysTyIleIleArgIleSerThrSerIleLeuAspLeuArg 800
|||||
2384 CAGCTCAAGATATATCATTCGAATAAGTACAAGTATCTTGATCTCAGA 2433
|||||
801 AspLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuIleProLy 817
|||||
2434 GACAAGTTCATGAATCTCTCAAGTGAATACTACTGCTCTCATCCAAA 2483
|||||
817 sGluAlaAsnSerGluGluValPheLeuPheLysProGluAsnIleThrp 834
|||||
2484 GGAAGCCAACTCTGAGGAAGTCTTTTGTGTTAAACCAAGAAACATTACTT 2533
|||||
834 heGluAsnGlyThrAspLeuPheIleAlaIleGlnAlaValAspLysVal 850
|||||
2534 TTGAAATGGCACAGATCTTTTCATTCCTATTTCAGGCTGTTGATAAGGTC 2583
|||||
851 AspLeuLysSerGluIleSerAsnIleAlaArgValSerLeuPheIlePr 867
|||||
2584 GATCTGAATTCAGAAATATATCAACATTCGACGAGATATCTTTGTATTTC 2633
|||||
867 oProGlnThrProProGluThrProSerProAspGluThrSerAlaProC 884
|||||
2634 TCCACAGACTCCCGCAGAGACACTAGTCTGATGAACGCTGCTGCTCTT 2683
|||||
884 ysProAsnIleHisIleAsnSerThrIleProGlyIleHisIleLeuLys 900
|||||
2684 GTCTAATATTCATATCAACAGACCATTCCTTGGCATTCACATTTTAAA 2733
|||||
901 IleMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914
|||||
2734 ATTATGTGAAGTGGATGATGAGAGAACTGCAGCTGTCAATAGCC 2775
|||||
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq.na2001.DAT:AAH46102

seq_documentation_block:

ID AAH46102 standard; DNA; 2742 BP.

XX AAH46102;

XX DT II-SEP-2001 (first entry)

XX DE Human CLCA1 coding sequence, SEQ ID NO:4.

XX KW Human CLCA1; goblet cell; mouse Gob-5 orthologue; drug screening;

XX KW expression inhibition; antisense therapy; gene therapy;

XX KW chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic;

XX CS

XX OS Homo sapiens.

XX FH Key

XX FT CD5 Location/Qualifiers

XX FT 1..2742

XX FT /*tag= a

XX FT /partial

XX FT /product= "Human CLCA1"

XX FT /note= "No stop codon given in the specification"

XX MO200138530-AL.

XX PD 31 MAY-2001.

XX PF 22-NOV-2000; 2000WO-JP08232.

XX PR 24-NOV-1999; 99JP-0333479.

XX PR 27-APR-2000; 2000JP-0127589.

XX PA

XX (TAKE) TAKEDA CHEM IND LTD.

XX PI

XX Nakanishi A, Morita S;

XX DR

XX WPI; 2001-355935/37.

XX DR P-PSDB; AAB73716.

XX PT New antisense nucleotide, useful for treatment and prevention of

XX PT bronchial asthma and chronic obstructive pulmonary disease

XX PS

XX Claim 3; Page 82-84; 104pp; Japanese.

XX CC The invention relates to an antisense nucleotide targetted to the mouse

XX CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,

XX CC the CLCA1 gene (coding sequence shown in AAH46102). The invention also

XX CC relates to an antibody specific for the Gob-5 protein, medical and

XX CC diagnostic compositions containing the antisense nucleotide or the

XX CC antibody, and methods and kits for screening for compounds which inhibit

XX CC the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.

XX CC The antisense oligonucleotides and antibody are therefore useful for the

XX CC treatment and prevention of bronchial asthma and chronic obstructive

XX CC pulmonary disease. The present sequence represents the human CLCA1

XX CC gene coding sequence.

XX SQ

XX Sequence 2742 BP; 833 A; 616 C; 623 G; 670 T; 0 other;

alignment_scores:

Quality: 4756.00 Length: 914

Ratio: 5.204 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.891

alignment_block:

US-09-049-696-41 x AAH46102

Align: seg 1/1 to: AAH46102 from: 1 to: 2742

MetClyProPheLysSerSerValPheIleLeuIleLeuHisLeuLeuG1 17

|||||

ATGGGGCCATTTAAGAGTCTGTGTTTCATCTTGATCTTCACCTTTCTAGA 50

17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34
 |||||
 51 AGGGGCCCTGAGTAATTCATCTATTGAGCTGAACAACAATGGCTATGAAG 100
 |||||
 34 lyIleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
 |||||
 101 GCATTGCTGTTGCAATCGACCCCAATGTGCCAGAAATGAAACACATTCATT 150
 |||||
 51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
 |||||
 151 CAACAAATAAAGGACATGTTGAGCCAGGCATCTCTATCTGTTTGAAGC 200
 |||||
 67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluT 84
 |||||
 201 TACAGAAAGCGATTTTATTTCAAAATGTTGCCATTTTGAATTCCTGAAA 250
 |||||
 84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
 |||||
 251 CATGGAAGACAAAGCTGACTATGTGAGACCAAAACTTGAGACCTTACAAA 300
 |||||
 101 AsnAlaAspValLeuValAlaGluSerThrProProGlyAsnAspGluPr 117
 |||||
 301 AATGCTGATGTTCTGCTGCTGACTCTACTCTCCAGGTAAATGATCAACC 350
 |||||
 117 oTyrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134
 |||||
 351 CTACACTGAGCAGATGGGCACTGTGGAGAGAGGGTGAAGGATCCACC 400
 |||||
 134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
 |||||
 401 TCACCTCTGATTTCAATGTCAGGAAAAAGTTAGCTGAATATGACCACAA 450
 |||||
 151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
 |||||
 451 GGTAAAGGCATTTGTCATGAGTGGGCTCATCTACGATGGGGAGTATTGA 500
 |||||
 167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGluA 184
 |||||
 501 CGAGTACATAATGATGAGAAATCTACTATCCAAATGGAAGAAATACAAG 550
 |||||
 184 laValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLysCys 200
 |||||
 551 CAGTAAGATGTTCCAGCAGGTATTACTGGTACAAATCTAGTAAAGAAGTGT 600
 |||||
 201 GlnGlyGlySerCysTyrThrLysArgCysThrPheAsnLysValThrGl 217
 |||||
 601 CAGGGAGGCAGCTGTTACACCAAAAGATGCACATTCAAATAAGTTACAGG 650
 |||||
 217 yLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrGluL 234
 |||||
 651 ACTCTATGAAAAGGATGTGAGTTGTTCTCCNATCCCGCCAGCAGGAGA 700
 |||||
 234 yAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPheCys 250
 |||||
 701 AGGCTTCTATATGTTGCACAACTGTTGATTCTATAGTTGAATTCGT 750
 |||||
 251 ThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLysCy 267
 |||||
 751 ACAGAAACAAACCCACAAAGAGCTCCAAACAGCAAAATCAAAAATG 800
 |||||
 267 sAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheLysL 284
 |||||
 801 CATCTCCGAAGCACATGGAGATGATCGTGATCTGAGGACTTTTAAAG 850
 |||||
 284 ysThrThrProMetThrThrGlnProProAsnProThrPheSerLeuLeu 300
 |||||
 851 AAACCACTCTATGACACACAGCCACCAAAATCCCACTTCTCATTTGCTG 900
 |||||
 301 GlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySerMe 317
 |||||
 901 CAGATTGGACAAAGAAATTGTTGTTAGTCTTGTGACAAATCTGGAAGCAT 950
 |||||

317 tAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheL 334
 |||||
 951 GCGGACTGTGTAACCGCTCAATCGACTGAATCAAGCAGGCCAGCTTTTCC 1000
 |||||
 334 euLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPheAsp 350
 |||||
 1001 TGCTGCAGACAGTTGAGCTGGGTCTCTGGTGGGATGGTGACATTTGAC 1050
 |||||
 351 SerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySerAs 367
 |||||
 1051 AGTGCTGCCCATGTACAAAAGTGAACATCATAGATAAACAGTGGCAGTGA 1100
 |||||
 367 parGaspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyThrS 384
 |||||
 1101 CAGGGACACACTCGCCAAAAGATTACTGCAGCAGCTTCAGGAGGACGT 1150
 |||||
 384 erIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLysTyr 400
 |||||
 1151 CCATCTGCAGCGGGCTTCGATCGGCATTTACTGTGATTAGGAAGAATAT 1200
 |||||
 401 ProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAsnTh 417
 |||||
 1201 CCAACTGATGGATCTGAATTTGTGCTGCTGACGGATGGGGAAGACAACAC 1250
 |||||
 417 rIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleHisT 434
 |||||
 1251 TATAAGTGGGTGCTTTAACGAGGTCAACAAAGTGGTGCATCATCCACA 1300
 |||||
 434 hrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluLeuSerLys 450
 |||||
 1301 CAGTCCGTTTGGGCGCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAA 1350
 |||||
 451 MetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnAsnGl 467
 |||||
 1351 ATGACAGGAGGTTTACAGACATATGTTTCAGATCAAGTTTCAGAACAAATGG 1400
 |||||
 467 yLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValSerC 484
 |||||
 1401 CCTCATGTATGCTTTTGGGGCCCTTTTCATCAGGAAATGGAGCTGTCTCTC 1450
 |||||
 484 InArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSerGln 500
 |||||
 1451 AGCGCTCCATCCAGCTTGAGAGTTAAGGGATTAAACCTCCAGAACACCCAG 1500
 |||||
 501 TrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspThrLe 517
 |||||
 1501 TGGATGAATGGCACAGTGTGCTGGACAGCCGCTGGGAAAGGACACTTT 1550
 |||||
 517 uPheLeuIleThrTrpThrThrGlnProGlnIleLeuLeuTrpAspP 534
 |||||
 1551 GTTCTTATACCTGGACAAACGAGCCTCCCAAAATCCTTCTCTGGGATC 1600
 |||||
 534 roSerGlyGlnLysGlnGlyPheValValAspLysAsnThrLysMet 550
 |||||
 1601 CCAGTGGACAGAAAGCAAGCTGGCTTTGTAGTGGACAAAACACCAAAATG 1650
 |||||
 551 AlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLysTyrSe 567
 |||||
 1651 GGCTACTCCAAATCCAGGCATTTGTAAGTTGGCACTTGGAAATATACAG 1700
 |||||
 567 rLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSerArgAlaS 584
 |||||
 1701 TCTGCAAGCAAGCTCACAAACCTTGACCTGACTGTACGTCCTCCGTCGT 1750
 |||||
 584 erAsnAlaThrLeuProProIleThrValThrSerLysThrAsnLysAsp 600
 |||||
 1751 CCAATGCTACCTGCTCCAAATTACAGTGACTTCCAAACGACAAAGGAC 1800
 |||||
 601 ThrSerLysPheProSerProLeuValValTyrAlaAsnIleArgGlnGl 617
 |||||
 1801 ACCAGCAAAATCCCGAGCCCTCTGTTAGTTTATGCAAAATATTCGCAAGG 1850
 |||||
 617 yAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerValA 634
 |||||

|||||
1851 AGCCTCCCAATTCACAGGCGAGTGTACAGCCCTGATTGAATCAGTGA 1900
634 sNglyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAlaAsp 650
|||||
1901 ATGGAAACAGTGTACCTTGGAACTACTGGATAATGGAGCAGGTGCTGAT 1950
651 AlaThrLysAspAspGlyValTyrSerArgTyrPheThrThrTyrAspTh 667
|||||
1951 GCTACTAAGGATGAGGCTGTCTACTCAAGGTATTTCAACAACCTATGACAC 2000
667 rAsnGlyArgTyrSerValLysValArgAlaLeuGlyGlyValAlaAla 684
|||||
2001 GAATGGTAGATACAGTGTAAAGTGGGGCTCTGGGAGAGCTTAACGCAG 2050
684 laArgArgArgValIleProGlnGlnSerGlyAlaLeuTyrIleProGly 700
2051 CCAGACGAGAGTGTATACCCAGCAGAGTGGAGCACTGTACATACCTGGC 2100
701 TrpIleGluAsnAspGluIleGlnTrpAsnProProArgProGluIleAs 717
|||||
2101 TGGATGGAATGATGAATACAAATGGAATCCCAAGACCTGAATTTAA 2150
717 nLysAspAspValGlnHisLysGlnValCysPheSerArgThrSerSerG 734
|||||
2151 TAAGGATGATGTTCAACACAAAGCAAGTGTGTTTCAGCAGAAACATCTCG 2200
734 lyGlySerPheValAlaSerAspValProAsnAlaProIleProAspLeu 750
2201 GAGGCTCATTTGTGGCTTCTGTATGCTCCCAATGCTCCATACCTGATCTC 2250
751 PheProProGlyGlnIleThrAspLeuLysAlaGluIleHisGlyGlySe 767
2251 TTCCACCTGCCAAATCACCAGCTGAAGCGGAAATTCACGGGGGCGAG 2300
767 rLeuIleAsnLeuThrTrpAlaProGlyAspAspTyrAspHisGlyT 784
2301 TCTCATTAACTGACTGGACAGCTCTCTGGGATGATTATGACCATCGAA 2350
784 hrAlaHisLysTyrIleIleArgIleSerThrSerIleLeuAspLeuArg 800
2351 CAGCTCACAACTATATCATTCGAATAGTACAAGTATTCTTGATCTCAGA 2400
801 AspLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuIleProLy 817
2401 CACAACTTCATGATATCTCTCAAGTGAATACTACTGCTCTCATCCCCAA 2450
817 sGluAlaAsnSerGluGluValPheLeuPheLysProGluAsnIleThrP 834
2451 GGAAGCCAACTCTGAGGAAGTCTTTTGTAAACCCAGAAACATTAATT 2500
834 heGluAsnGlyThrAspLeuPheIleAlaIleGlnAlaValAspLysVal 850
2501 TTGAAATGGCACAGATCTTTTCAATGCTATTCAGGCTGTGTAAGGTC 2550
851 AspLeuLysSerGluIleSerAsnIleAlaArgValSerLeuPheIlePr 867
2551 CATCTGAAATCAGAAATATCCAAATGACACAGTATCTTTGTTTATTC 2600
867 oProGlnThrProProGluThrProSerProAspGluThrSerAlaProC 884
2601 TCCACAGACTCCGCCAGACACACCTAGTCTCTGATGAAGCTGCTGCTCT 2650
884 ysProAsnIleHisIleAsnSerThrIleProGlyIleHisIleLeuLys 900
2651 GTCTTAATATTATCATATCAACAGCACCATTCTCTGCATTACATTTTAAA 2700
901 ileMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914
2701 ATTATGTGAAGTGGATAGGAGAACTCAGCTGTCAATGACC 2742

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AA129502

seq_documentation_block:
ID_AA129502 standard; cDNA; 3311 BP.
XX
AC_AA129502:
XX
DT_12-OCT-2001 (first entry)
XX
DE_C502P determined cDNA sequence.
XX
KW_Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW_immunogenic; gene therapy; vaccine; colonic cancer; ss.
XX
OS_Homo sapiens.
XX
PN_W00000149716-A2.
XX
PD_12 JUL -2001.
XX
PF_29-DEC-2000; 2000MO-US35596.
XX
PR_30-DEC-1999; 99US-0476296.
PR_10-JAN-2000; 2000US-0480321.
PR_15-FEB-2000; 2000US-0504629.
PR_06-MAR-2000; 2000US-0519444.
PR_19-MAY-2000; 2000US-0575251.
PR_25-JUN-2000; 2000US-0609448.
PR_28-AUG-2000; 2000US-0649811.
XX
PA_(CORI-) CORIXA CORP.
XX
PI_Xu J. Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
FI_King GE, Wang T, Jiang Y;
XX
DR_WPI; 2001-441847/47.
XX
PT_Colon tumor associated proteins and nucleic acids useful for the
XX_prevention, diagnosis and treatment of colonic cancer
XX
PS_Claim 2; Page 425-426; 472pp; English.
XX
CC_The present invention describes colon tumour associated proteins (I) and
CC_the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC_(I) and (II) can be used in gene therapy and vaccine production. (I) and
CC_(II) may be used in the prevention, diagnosis and treatment of diseases
CC_associated with inappropriate colon tumour associated protein (TCAP)
CC_expression, such as colonic cancer. For example, (I) and (II) may be
CC_used to treat disorders associated with decreased expression by
CC_rectifying mutations or deletions in a patient's genome that affect the
CC_activity of TCAPs by expressing inactive proteins or to supplement the
CC_patients own production of them. Additionally, (II) may be used to
CC_produce the TCAP proteins, by inserting the nucleic acids into a host
CC_cell culturing the cell to express the protein. (II) and its
CC_complementary sequences may also be used as DNA probes in diagnostic
CC_polymerase chain reaction (PCR) and hybridisation assays to detect and
CC_quantitate the presence of similar nucleic acids in samples, and
CC_therefore which patients may be in need of restorative therapy. (I) may
CC_also be used as antigens in the production of antibodies against TCAPs
CC_and in assays to identify modulators of TCAP expression and activity.
CC_Anti-(I) antibodies and antagonists may also be used to down regulate
CC_TCAP expression and activity. The anti-(I) antibodies may also be used
CC_as diagnostic agents for detecting the presence of TCAPs in samples
CC_(e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512
CC_and AA24494 to AA24523 represent nucleotide and amino acid sequences
CC_given in the exemplification of the present invention.
XX
SQ_Sequence 3311 BP; 1028 A; 692 C; 742 G; 849 T; 0 other;

alignment_scores:
Quality: 4756.00 Length: 914
Ratio: 5.204 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.891

alignment_block:

US-09-049-696-41 x AAI29502

Align seq 1/1 to: AAI29502 from: 1 to: 3311

```
1 MetGlyProPheLysSerValPheIleLeuIleLeuHisLeuLeuG1 17
352 ATGGGGCCATTAAAGAGTCTGTGTTTCATCTTCTACCTTCTAGA 401
17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrgluG 34
402 AGGGGCCCTGAGTAATTCATCTATTGAGCTGAACAACATGGCTATGAAG 451
34 lytleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
452 GCATTGTCGTTGCAATCGACCCCAATGTGCCAAGATGAACACATCATT 501
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
502 CAACAAATAAAGCACATGGTGACCCAGGCATCTCTGTATCTGTTGAAGC 551
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluT 84
552 TACAGGAAGACGATTTTATTTCACAAATGTTGCCATTTTTCATCTCTGAAA 601
84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
602 CATGGAGACAAAGCTGACTATGTGAGACCAAACTTGAGACCTACAAA 651
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
652 AATGCTCATCTTCTGTTGCTGAGTCTACTCTCCAGGTAATGATGAACC 701
117 oTyrThrGluGlnMetGlyAsnGlyGluLysGlyGluArgIleHisL 134
702 CTACACTGACAGATGGCAACTGTGAGAGAAAGGTTGAAGGATCCACC 751
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
752 TCACCTCTGATTTTCATGTGAGGAAAAGTTAGCTGAATATGACCACAA 801
151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
802 GGTAAAGGCATTTCTCCATGAGTGGGCTCATCTACGATGGGAGCTATTGA 851
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGlnA 184
852 CGAGTACAATAATGATGAGAAATTCCTACTTATCCAAATGGAAGAATAACAAG 901
184 laValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLysCys 200
902 CAGTAGATGTTCCAGCAGGTATTACTGGTACAAATGTAGTAAAGAGTGT 951
201 GlnGlyCysSerCysTyrThrLysArgCysThrPheAsnLysValThrG1 217
952 CAGGGAGGCAGCTGTACACCAAAAGATGCACATTCAATAAAGCTTACAGG 1001
217 yLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrGluL 234
1002 ACTCTATGAAAAAGGATGTGAGTTCCTTCCCAATCCCGCCAGCAGGAGA 1051
234 ysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPheCys 250
1052 AGCCTTCTATATGTTTGCACACATGTTGATTTCTATAGTTGAATCTGT 1101
251 ThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLysCy 267
1102 ACAGACAAAAACCAACAAGAAAGCTCCAAACAAGCAAAATCAAAAATG 1151
267 sAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheLysL 284
1152 CAATCTCCGAAGCACATGGGAAGTGTGCTGATTCTGTGAGGACTTTAAGA 1201
```

```
284 ysThrThrProMetThrThrGlnProProAsnProThrPheSerLeuLeu 300
1202 AAACCACTCTCTATGACAAACACACGCCAACATCCACCTTCTCATGTGCTG 1251
301 GlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySerMe 317
1252 CAGATTGGACAAAGAAATTGTGTGTTTAGTCTTGACAAATCTGGAACCAT 1301
317 tAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheL 334
1302 GGGGACTGGTTAACCGCTCAATCGACTGAATCAAGCAGGCGACGCTTTCC 1351
334 euLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPheAsp 350
1352 TGTGTCAGACAGTTGAGCTGGGCTCTGGGTGGGATGGTGACATTGAC 1401
351 SerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySerAs 367
1402 AGTGCTGCCCATGTACAAAGTGAACCTACATACAGATAACAGTGGCAGTGA 1451
367 pArgAspThrLeuAlaLysArgLeuProAlaAlaSerGlyGlyThrS 384
1452 CAGGGACACACTGCCCAAAAGATTAGCTGCAGCAGCTTCAGGAGGACGT 1501
384 erIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLysTyr 400
1502 CCATCTGCAGCGGGCTTCGATCGGCATTTTACTGTGATTAGGAAGAATAT 1551
401 ProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAsnTh 417
1552 CCAACTGATGGATCTGAAATTTGTGCTGCTGACGGATGGGGAACACACAC 1601
417 rIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleHist 434
1602 TATAAGTGGTGCTTTTAAACGAGGTCAACAAAGTGTGTCATCATCCACA 1651
434 hrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluLeuSerLys 450
1652 CAGTCCCTTTGGGGCCCTCTGCAGCTCAAGACTAGAGGAGCTGTCCAAA 1701
451 MetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnAsnG1 467
1702 ATCAGCAGGAGTTTACAGACATATGCTTCAGATCAAGTTCAGAAACAATGG 1751
467 yLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValSerG 484
1752 CCTCATGTGATGCTTTTGGGGCCCTTTTCATCAGGAAATGGAGCTGTCTCTC 1801
484 InArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSerGln 500
1802 AGCGCTCCATCCAGCTTGAGTAAAGGATTAACCTCCAGAACACGCCAG 1851
501 TrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspThrLe 517
1852 TGGATCAATGGCACACGTGATCGTGGACAGCACCGCTGGGAAAGCACACTTT 1901
517 uPheLeuIleThrTrpThrThrGlnProProGlnIleLeuLeuTrpAspP 534
1902 GTTTCCTATCACCTGGACACAGCAGCTCCCAAAATCCTTCTCTGGGATC 1951
534 roSerGlyGlnLysGlnGlyPheValValAspLysAsnThrLysMet 550
1952 CCAGTGGACAGAAAGAGGTGGCTTTGTAGTGGACAAACACCAAAATG 2001
551 AlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLysTyrSe 567
2002 GCCTACTCCAAATCCAGGCATTGCTAAGTTGGCATTGGAAATACAG 2051
567 rLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSerArgAlaS 584
2052 TCTGCAAGCAAGCTCACAAACCTTGACCTGACTGTACGTCCTCCGTCGT 2101
584 erAsnAlaThrLeuProProIleThrValThrSerLysThrAsnLysAsp 600
```

```
|||||
2102 CCAATGCTACCTCGCTCCCAATTACAGTGACTTCCAAAACGAACAGGAC 2151
601 ThrSerLysPheProSerProLeuValValTyrAlaAsnIleAtrGlnG1 617
|||||
2152 ACCAGCAAAATCCCCAGCCCTCTGGTACTTTATGCAAAATATTCGCCAAG 2201
617 yAlaSerProIleLeuAtrGAlaSerValThrAlaLeuIleGluSerValA 634
|||||
2202 AGCTTCCCAATCTCAGGGCCAGTGTACAGCCCTGATTGAATCAGTGA 2251
634 snGlyLysThrValThrLeuGluLeuLeuAspAsnGlyValAclyAlaAsp 650
|||||
2252 ATGGAAAAACAGTTACCTTGAACCTACTGGATAATGGAGCAGGTGCTGAT 2301
651 AlaThrLysAspAspGlyValTyrSerArgTyrPheThrThrTyrAspTh 667
|||||
2302 GCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCAACAACCTTATGACAC 2351
667 rAsnGlyArgTyrSerValLysValArgAlaLeuGlyGlyValAlaAsnAlaA 684
|||||
2352 GAATGGTAGATACAGTGTAAAGTCGGGCTCTGGAGAGGTTAACGCAG 2401
684 laArgArgArgValIleProGlnGlnSerGlyAlaLeuTyrIleProGly 700
|||||
2402 CCAGACGAGAGTGTATACCCAGCAGAGTGGAGCACTGTACATACCTGGC 2451
701 TrpIleGluAsnAspGluIleGlnTrpAsnProProArgProGluIleAs 717
|||||
2452 TGGATTGAGATGATGAATAACAATCAATGAATCCACCAACACCTGAAATTA 2501
717 nLysAspValGlnHisLysGlnValLysPheSerArgThrSerG 734
|||||
2502 TAAGGATGATGTCAACACAGCAAGTGTGTTACAGCAGAAATCCTCGG 2551
734 lyGlySerPheValAlaSerAspValProAsnAlaProIleProAspLeu 750
|||||
2552 GAGGCTCATTTGTGGCTTCTGATGTCCCAATATGCTCCCATACCTGATCTC 2601
751 PheProGlyGlnIleThrAspLeuLysAlaGluIleHisGlyGlyse 767
|||||
2602 TTCCACCTGGCCAAATCACCACCTGAAGCGGAAATTCACGGGGCGAG 2651
767 rLeuIleAsnLeuThrTrpThrAlaProGlyAspAspTyrAspHisGlyT 784
|||||
2652 TCTCATTAATCTGACTGGACAGCTCCTGGGGATGATTATGACCATGAA 2701
784 hrAlaHisLysTyrIleIleArgIleSerThrSerIleLeuAspLeuArg 800
|||||
2702 CAGCTCACAAAGTATATCATTCGAATAAGTACAAAGTATTTCTTGATCTCAGA 2751
801 AspLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuIleProLy 817
|||||
2752 GACAAGTTCAATGAATCTTCAAGTGAATACTACTGCTCTCATCCCAA 2801
817 sCluAlaAsnSerGluGluValPheLeuPheLysProGluAsnIleThrP 834
|||||
2802 GGAAGCCCAACTCTGAGGAAGTCTTTTCTTTAAACAGAAAACATTAATT 2851
834 heGluAsnGlyThrAspLeuPheIleAlaIleGlnAlaValAspLysVal 850
|||||
2852 TTGAAAATGGCACAGATCTTTTTCATTGCTATTCAGGCTGTTCATAAGGTC 2901
851 AspLeuLysSerGluIleSerAsnIleAlaArgValSerLeuPheIlePr 867
|||||
2902 GATCTGAAATCAGAAATATCCAACATTCACAGATATCTTTGTTTATTTCC 2951
867 oProGlnThrProProGluThrProSerProAspGluThrSerAlaProC 884
|||||
2952 TCCACAGACTCCGCCAGACACCTAGTCCCTGATGAACAGCTGCTGCTCT 3001
884 ysProAsnIleHisIleAsnSerThrIleProGlyIleHisIleLeuLys 900
|||||
```

```
3062 GTCCTAATATTTCATATCAACAGCACCATTCTCGCATTCACATTTAAAA 3051
901 IleMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914
|||||
3052 ATTATGTGGAAGTGTAGTAGGAGAACTGCAGCTGTCAATAGCC 3093
```

seq_name: /SDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH33285

seq_documentation_block:

ID AAH33285 standard; cDNA; 2867 BP.

XX

AC AAH33285;

XX

DF 03-SEP-2001 (first entry)

XX

DE Human colon cancer antigen encoding cDNA SEQ ID NO:341.

XX

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;

KW colorectal carcinoma; Chromosome 1; ss.

XX

CS Homo sapiens.

XX

PN W0200122920-A2.

XX

PD 05-APR-2001.

XX

PF 28-SEP-2000; 2000WO-US26524.

XX

PR 29-SEP-1999; 99US-0157137.

PR

03-NOV-1999; 99US-0163280.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX

DR WPI; 2001-235357/24.

DR

P-PSDB; AAG73854.

XX

FT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,

PT useful for preventing, diagnosing and/or treating colorectal cancers -

XX

PS Clinim 1; Page 2452-2453; 9803pp; English.

XX

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

CC cancer-associated nucleic acid molecules (N) and proteins (P), where

CC the proteins are collectively known as colon cancer antigens. The colon

CC cancer antigens have cytostatic activity and can be used in gene

CC therapy and vaccine production. N and P may be used in the prevention,

CC diagnosis and treatment of diseases associated with inappropriate P

CC expression. For example, N and P may be used to treat disorders

CC associated with decreased expression by rectifying mutations or deletions

CC in a patient's genome that affect the activity of P by expressing

CC inactive proteins or to supplement the patients own production of P.

CC Additionally, N may be used to produce the colon cancer-associated Ps,

CC by inserting the nucleic acids into a host cell and culturing the cell

CC to express the proteins. N and P can be used in the prevention, diagnosis

CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204

CC and AAB77789 represent sequences used in the exemplification of the

CC present invention.

CC N.A. Pages 666 to 682 and page 7053 of the sequence listing were

CC missing at time of publication, meaning no sequences are present for

CC SEQ ID NO:1027 to 1052, 7921 and 7922.

alignment_scores:

Quality: 4740.00 Length: 915

Ratio: 5.186 Gaps: 1

Percent Similarity: 99.891 Percent Identity: 99.781

alignment_block:

US-09-049-696-41 x AAH33285

Sequence 2867 BP; 891 A; 633 C; 643 G; 700 T; 0 other;

Align seg 1/1 to: AAH33285 from: 1 to: 2867

1 MetGlyProPheLysSerValPheIleLeuIleLeuHisLeuLeuG1 17
|||||
38 ATGGGGCCATTTAAGAGTCTGTGTTCACTTGATCTTCACCTTCTAGA 87
|||||
17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34
|||||
88 AGGGGCCCTGAGTAATTCATCTATTGAGTGAACAACAATGGCTATGAAG 137
|||||
34 lyIleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
|||||
138 GCATTGTGTTGCAATCGACCCCAATGTGCCAGAAGATGAACACACTATT 187
|||||
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
|||||
188 CACAANAAGGACATGTGAGCCAGGCATCTCTGTATCTGTTGAAGC 237
|||||
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluT 84
|||||
238 TACAGGAAGCGGATTTATTTCAAAAATGTTGCCAATTTGATTCCTGAAA 287
|||||
84 hTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
|||||
288 CATGGAAGACAAAGGCTGACTATGTGAGACCAAACTTGAGACCTACAAA 337
|||||
101 AsnAlaAspValLeuValAlaGluSerThrProProGlyAsnAspGluPr 117
|||||
338 AATGCTGATGTTCTGCTGCTGAGTCTACTCTCCAGGTAAATGATGACC 387
|||||
117 oTyrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134
|||||
388 CTACACTGAGCAGATGGCAACTGTGGAGAAAGGTTGAAGGATCCACC 437
|||||
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
|||||
438 TCACCTCTGATTTCTGAGGAAAGATTAGCTGAATGAGCCACAA 487
|||||
151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
|||||
488 GGTAGGCAATTTCTCCATGAGTGGGCTCATCTACGATGGGAGTATTGA 537
|||||
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGlnA 184
|||||
538 CGAGTACAATAATGATGAGAAATTCATCTATCCAAATGGAAGAAATACAAG 587
|||||
184 laValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLysCys 200
|||||
588 CAGTAAGATCTTCAGCAGGTATTACTGGTACAAATGTAGTAAGAAGTGT 637
|||||
201 GlnGlyCysSerCysTyrThrLysArgCysThrPheAsnLysValThrG1 217
|||||
638 CAGGGAGGCAAGCTGTTACACCAAAAGATGCACATTCAATAAAGTTACAGG 687
|||||
217 yLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrGluL 234
|||||
688 ACTCTATGAAAAGGATGTGAGTGTGTTCTCCAAATCCCGCCAGCAGGAGA 737
|||||
234 ysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPheCys 250
|||||
738 AGGCTTCTATAATGTTTGACACACATGTTGATCTATAGTTGAATCTGT 787
|||||
251 ThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLysCy 267
|||||
788 ACAGAACAAAACCAACAAGAAGCTCCAAACAAGCAAAATCAAAAATG 837
|||||
267 sAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheLysL 284
|||||
838 CAATCTCCGAAGCAGATGGGAAGTGATCCGTGATCTTGAGGACTTTAAGA 887
|||||
284 ystThrProMetThrThrGlnProAsnProThrPheSerLeuLeu 300
|||||

888 AAACCACTCCTATGACAAACACAGCCCAAAATCCCACTTCTCATTTGCTG 937
|||||
301 GlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySerMe 317
|||||
938 CAGATTGGACAAAGAAATTTGTGTTTAGTCTTGACAAATCTGGAAGCAT 987
|||||
317 tAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheL 334
|||||
988 GCGGACTGTGTAACCGCTCAATCGACTGAATCAAGCAGGCCAGCTTTTCC 1037
|||||
334 euLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPheAsp 350
|||||
1038 TGTCTGCAGACAGTTGAGCTGGGTCTGGTGGGATGGTGACATTTGAC 1087
|||||
351 SerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySerA 367
|||||
1088 AGTGTGCCCCATGTACAAAGTGAACATCATACAGATAAACAGTGGGCAGTG 1137
|||||
367 sPArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyThr 383
|||||
1138 ACAGGGACACACTCGCCAAAGATTACCTGCAGCAGCTTCAGGAGGAGC 1187
|||||
384 SerIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLysTy 400
|||||
1188 TCATCTGCAGCGGGCTTCGATCGGCATTTACTGTGATTAGNAGAAATA 1237
|||||
400 rProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAsnT 417
|||||
1238 TCCAACGTGATGGATCTCAAAATTTGCTGCTGACGGATGGGAAGACAACA 1287
|||||
417 hrIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleHis 433
|||||
1288 CTATAAGTGGGTGCTTTAAACGAGGTCAAAACAAAGTGTGCCATCATCCAC 1337
|||||
434 ThrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluGluLeuSerLy 450
|||||
1338 ACAGTCGCTTTGGGGCCCTCTCGAGCTCAAGAACTAGAGGAGCTGTCCA 1387
|||||
450 sMetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnAsnG 467
|||||
1388 AATGACAGAGGTTTACAGACATATGCTTCAGATAAGTTTCAGAACAAATG 1437
|||||
467 lyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValSer 483
|||||
1438 GCCTCATGATGCTTTTGGGGCCCTTTCATCAGAAATGGAGCTGTCTCT 1487
|||||
484 GlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSerGl 500
|||||
1488 CAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAACCTCCAGAACACGCCA 1537
|||||
500 nTrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspThrL 517
|||||
1538 GTGGATGAATGGCACAGTATCGTGGACAGCACCGTGGGAAAGGACACTT 1587
|||||
517 euPheLeuIleThrTrpThrThrGlnProProGlnIleLeuLeuTrpAsp 533
|||||
1588 TGTTTCTTATCACCCTGACAAACGACCCCTCCCAAAATCCTTCTCTGGAT 1637
|||||
534 ProSerGlyGlnLysGlnGlyGlyPheValValAspLysAsnThrLysMe 550
|||||
1638 CCCAGTGGACAGAAGCAAGGTGGCTTTGTAGTGGACAAAAAACACCAAAAT 1687
|||||
550 tAlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLysTyrS 567
|||||
1688 GGCTTACCTCCAAATCCAGGCATTCCTAAGTTGGCACTTGGAAATACA 1737
|||||
567 erLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSerArgAla 583
|||||
1738 GTCTGCAACCAAGCTCACAACCTTGACCTGACTGTACAGTCCCGTGGC 1787
|||||
584 SerAsnAlaThrLeuProProIleThrValThrSerLysThrAsnLysAs 600
|||||
1788 TCCAATGCTACCTGCCTCCATTTACAGTACTTCCAAAACGAACAAGGA 1837
|||||

```
600 pThrSerLysPheProSerProLeuValValThrAlaAsnIleArgGlnG 617
1838 CACCAGCAAAATCCCGAGCCCTCTGGTAGTTTATGCAAAATATTCGCCAAG 1887
617 lYAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerVal 633
1888 GAGCCTCCCAATTTCTCAGGGCCAGTGTACAGCCCTGATGATGAATCAGTG 1937
634 AsnGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAlaAs 650
1938 AATGAAAAACAGTTACCTTGCAACTACTGCAATAATGGAGCAGGTGCTGA 1987
650 pAlaThrLysAspAspGlyValThrSerArgTyrPheThrThrTyrAspT 667
1988 TGCTACTAAGGATCAGCGTGTCTACTCAAGGTATTTCAACAACCTTATGACA 2037
667 hrAsnGlyArgTyrSerValLysValArgAlaLeuGlyGlyValAsnAla 683
2038 CCAATGGTAGATACAGTGTAAAGTGGGGCTCTGGGAGGAGTTAACGCA 2087
684 AlaArgArgValIleProGlnGlnSerGlyAlaLeuTyrIleProGln 700
2088 GCCAGCGGAGTGTATACCCAGCAGAGTGGAGCCTGTACATACCTGG 2137
700 yTrpIleGluAsnAspGluIleGlnTyrAsnProArgProGluIleAla 717
2138 CTGATTTAGAAATGATGAATACAAATGGAATCCCAAGACCTGAATTA 2187
717 snLysAspAspValGlnHisLysGlnValCysPheSerArgThrSerSer 733
2188 ATAAGGATGATGTTCACACAAAGCAAGTGTCTTCAGCAGAACATCTCTCG 2237
734 GlyGlySerPheValAlaSerAspValProAsnAlaProIleProAspLe 750
2238 CGAGGCTCATTTGGTGTCTCTGATGTCCTCCAAATGCTCCATACCTGATCT 2287
750 uPheProProGlyGlnIleThrAspLeuLysAlaGluIleHisGlyGlyS 767
2288 CTTCACCTCGCCAAATCACCGACCTGAAGCGGAAATTCACGGGGGCA 2337
767 erLeuIleAsnLeuThrTrpThrAlaProGlyAspAspTyrAspHisGly 783
2338 CTCTCATTAATCTACCTTGGACAGCTCTCTGGGATGATTATGACCATGGA 2387
784 ThrAlaHisLysTyrIleIleArgIleSerThrSerIleLeuAspLeuAr 800
2388 ACAGCTCAAGTATATCATTCGAATAAGTACAAGTATCTTGATCTCAG 2437
800 gAspLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuIleProL 817
2438 AGACAAGTTCATGAATCTCTCAAGTGAATACTACTGCTCTCATPCCCA 2487
817 ysGluAlaAsnSerGluGluValPheLeuPheLysProGluAsnIleThr 833
2488 AGGAAGCAACTCTCAGGAAGTCTTTTGTCTTAAACACAGAACCATTA 2537
834 PheGluAsnGlyThrAspLeuPheIleAlaIleGlnAlaValAspLysVa 850
2538 TTTGAAAATGGCAGACATCTTTTCAATTCCTATTCAGGCTGTTGATAAGT 2587
850 lAspLeuLysSerGluIleSerAsnIleAlaArgValSerLeuPheIleP 867
2588 CGATCTGAATCAGAAATATCAACATTCGACGAGTATCTTTGTTTATTC 2637
867 roProGlnThrProProGluThrProSerProAspGluThrSerAlaPro 883
2638 CTCCACAGACTCCGCCAGACACCTAGTCTCTGATGAACGCTGCTGCTCT 2687
884 CysProAsnIleHisIleAsnSerThrIleProGlyIleHisIleLeuLy 900
2688 TGTCTTAATATTCATATCAACAGCACCATTCTTGGCATTCACATTTTAA 2737
```

```
900 silemettrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914
2738 AATTATGTGGAAGTGGATAGGAGAACTGCAGCTGTCAATAGCC 2780
seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAH35019
```

seq_documentation_block:

ID AAH35019 standard; cDNA; 3109 BP.

XX AAH35019;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen encoding cDNA SEQ ID NO:2101.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

XX colorectal carcinoma; ss.

XX Homo sapiens.

XX WO20012920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

XX 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI: 2001-235357/24.

XX P-PSDB; AAG75614.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,

XX useful for preventing, diagnosing and/or treating colorectal cancers -

XX Claim 1; Page 3587-3588; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patients own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent sequences used in the exemplification of the present invention.

XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 3109 BP; 970 A; 652 C; 687 G; 797 T; 3 other;

alignment_scores:

Quality: 4476.00 Length: 862

Ratio: 5.211 Gaps: 0

Percent Similarity: 99.652 Percent Identity: 99.536

alignment_block:

US-09-049-696-41 x AAH35019

Align Seg 1/1 to: AAH35019 from: 1 to: 3109

53 IleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAlaThrGI 69
|||||
44 ATTCGGCAGCAGGTGACCCASGCATCTCTGTATCTCTTTGAAGCTACAGG 93

69 YLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluThrTrpL 86
|||||
94 AAAGCGATTTATTTCAAAATGTTGCCATTTTGGATTCCTGAACCATGGA 143

86 YsthrLysAlaAspTyrValArgProLysLeuGluThrTyrLysAsnAla 102
|||||
144 AGACAAGGCTGACTATGTGAGACCAAACTTGAGACCTACAAAAATGCT 193

103 AspValLeuValAlaGluSerThrProGlyAsnAspGluProTyrTh 119
|||||
194 GATGTTCTGTTGCTGAGTCTACTCTCCAGGTAATGATGAACCCCTACAC 243

119 rGluGlnMetGlyAsnCysGlyLysGlyGluArgIleHisLeuThrP 136
|||||
244 TGAGCAGATGGGCAACTGTGGAGAGAGGTGAAGGATCCACCTCACCTC 293

136 roAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGlnGlyArg 152
|||||
294 CTGATTTTCATTCAGCAAAAAAGTTAGCTGAATATGACCACCAAGGTAGG 343

153 AlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAspGluTy 169
|||||
344 GCATTTTGTCCATGAGTGGGCTCATCTACGATGGGGAGTATTTGACAGTA 393

169 rAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGlnAlaVala 186
|||||
394 CATAATGATGAGAAATTTACTATTCCATGGGAAGAAATACAGCAGTAA 443

186 rgCysSerAlaGlyIleThrGlyThrAsnValValLysLysCysGlnGly 202
|||||
444 GATGTTTCAGCAGGTATTACTGGTACAAATGTAGTAAAGAAGTGTACAGGA 493

203 GlySerCysTyrThrLysArgCysThrPheAsnLysValThrGlyLeuTy 219
|||||
494 GGCAGCTGTTTACACCAAAAGATGCATTTCAATAAAGTACAGGACTCTA 543

219 rGluLysGlyCysGluPheValLeuGlnSerArgGlnThrGluLysAlas 236
|||||
544 TGAANAAGGATGTGAGTTGTTCTCCATCCCGCCAGCGGAGAGGCTT 593

236 erileMetPheAlaGlnHisValAspSerIleValGluPheCysThrGlu 252
|||||
594 CTATAATGTTTGACAAACATGTTGATTTCTATAGTTGAAATTTCTGTACAGAA 643

253 GlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLysCysAsnLe 269
|||||
644 CAAACCCACAAACAAGAGCTCCAAACAAGCAAAATCAAAATATGCAATCT 693

269 uArgSerThrTrpGluValIleArgAspSerGluAspPheLysLysThrT 286
|||||
694 CCGAAGCACATGGGAAGTGCATCGTGATTTCTGAGGACTTTTAAGAAACCA 743

286 hrProMetThrThrGlnProProAsnProThrPheSerLeuLeuGlnIle 302
|||||
744 CTCCTATGACACACACGCCACCAAAATCCCACTTCTCATTTGCTGCGAGATT 793

303 GlyGlnArgIleValCysLeuValLeuAspLysSerGlySerMetAlaTh 319
|||||
794 GGACAAGAATTGTGTGTTTGTAGTTCCTTGACAAATCTGGAAGCATGGCGAC 843

319 rGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheLeuLeuG 336
|||||
844 TGTAAACCGGCTCAATCGACTGAATCAAGCAGGCCAGCTTTTCTGCTGC 893

336 lnThrValGluLeuGlySerTrpValIcylMetValThrPheAspSerAla 352
|||||
894 AGACAGTTGAGCTGGGGTCTCTGGGTGGGATGTGTGACATTTGACAGTGCT 943

353 AlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySerAspArgAs 369
|||||
944 GCCCATGTACAAAGTGAACATCATACAGATAAACAGTGGCAGTCACAGGA 993

369 pThrLeuAlaLysArgLeuProAlaAlaSerGlyGlyThrSerIleC 386
|||||
994 CACACTCGCCAAAGATTTACCTGCAGCAGCTTCAGGAGGAGCTCCATCT 1043

386 ysSerGlyLeuArgSerAlaPheThrValIleArgLysLysTyrProThr 402
|||||
1044 GCAGCGGCTTCGATCGGCATTTACTGTGATTAGGAAGAAATATCCAAC 1093

403 AspGlySerGluIleValLeuLeuThrAspGlyGluAspAsnThrIleSe 419
|||||
1094 GATGGATCTGAAATTTGTGCTGACGGATGGGGAAGACAACACTATAAG 1143

419 rGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleHisThrVala 436
|||||
1144 TGGGTGCTTTAAGAGGTCAACAAAGTGTGTCATCATCCACACAGTCCG 1193

436 laLeuGlyProSerAlaAlaGlnGluLeuGluLeuSerLysMetThr 452
|||||
1194 CTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAAATGACA 1243

453 GlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnAsnGlyLeuIl 469
|||||
1244 GGAGGTTTTACAGACATATGCTTCAGATCAAGTTTCAGAACAAATGGCCCTCAT 1293

469 eaSpAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValSerGlnArgS 486
|||||
1294 TGATGCTTTTGGGCCCTTTTCATCAGAAATGGAGCTGTCTCTCAGCGCT 1343

486 erileGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSerGlnTrpMet 502
|||||
1344 CCATCCAGCTTGAGAGTAAGGGATTAAACCTCCAGAACAGCCAGTGGATG 1393

503 AsnGlyThrValIleValAspSerThrValGlyLysAspThrLeuPheLe 519
|||||
1394 AATGGCACAGTGTATCGTGACAGCACCGCTGGGAAAGGACACTTTGTTCT 1443

519 utleThrTrpThrGlnProGlnIleLeuLeuTrpAspProSerG 536
|||||
1444 TATCACCCTGGACAAACGACGCTCCCAAAATCTCTCTGGGATCCCAAGT 1493

536 lysGlnLysGlnGlyGlyPheValValAspLysAsnThrLysMetAlaTyr 552
|||||
1494 GACAGAAGCAAGGTGGCTTTGTAGTGACAAACACCAAAATGGCCTAC 1543

553 LeuGlnIleProGlyIleAlaLysValGlyThrTrpLysTyrSerLeuGI 569
|||||
1544 CTCCAATCCCGAGCATTCCTAAGTTGGCACTTGGAAATACAGTCTGCA 1593

569 nAlaSerSerGlnThrLeuThrLeuThrValThrSerArgAlaSerAsnA 586
|||||
1594 AGCAAGCTCACAAACCTTGACCTGACTGTACGTCCTCCGCTGGCTCCAATG 1643

586 laThrLeuProPheThrValThrSerLysThrAsnLysAspThrSer 602
|||||
1644 CTACCTGCTCTCAATTCAGTACTTCCAAACGAAACAGGAGGACCCAGC 1693

603 LysPheProSerProLeuValValTyrAlaAsnIleArgGlnGlyAlase 619
|||||
1694 AAATTCGCCAGCCCTCTGGTAGTTTATGCAAAATATTCGCCAAGGACCTC 1743

619 rProIleLeuArgAlaSerValThrAlaIleLeuGluSerValAsnGlyL 636
|||||
1744 CCCAATTCCTCAGGGCCAGTGTACAGCCCTGATTGAATCAGTGAATGGA 1793

636 ysthrValThrLeuGluLeuAspAsnGlyAlaGlyAlaAspAlaThr 652
|||||
1794 AAACAGTTTACCTTGGAACTACTGGATTAATGGACAGGTGCTGATGCTACT 1843

653 LysAspAspGlyValTyrSerArgTyrPheThrThrTyrAspThrAsnGI 669


```
1844 AAGGATGACGGTGTCTACTCAGGTATTTCAACACTTATGACACGAATGG 1893
669 YAGTyrSerValLysValArgAlaLeuGlyGlyValAsnAlaLaArgA 686
1894 TAGATACAGTGTAAAGTGGCGGTCTGGGAGGAGTTAACGACGACGAC 1943
686 rGArgValIleProGlnGlnSerGlyAlaLeuTyrIleProGlyTrpIle 702
1944 GGAGAGTATACCCCGACAGAGTGGAGCACTGTACATACCTGGCTGGATT 1993
703 GluAsnAspGluIleGlnTrpAsnProArgProGluIleAsnLysAs 719
1994 GAGATGATGAATACATAGTATCCACCAAGACCTGAATTAATAGGA 2043
719 pAspValGlnHisLysGlnValCysPheSerArgThrSerSerGlyGly 736
2044 TGATGTTCAACACGAAGCAAGTGTGTTTCAGCAGAACATCTCTGGGAGCT 2093
736 erPheValAlaSerAspValProAsnAlaProIleProAspLeuPhePro 752
2094 CATTTGTGGCTTCTGATGTCCAAATGCTCCATACCTGATCTCTCCCA 2143
753 ProGlyGlnIleThrAspLeuLysAlaGluIleHisGlyGlySerLeuI 769
2144 CTGCGCAAAATCACGACCTGAGCGGAATTCACGGGGGCGAGTCTCAT 2193
769 eAsnLeuThrTrpThrAlaProGlyAspAspTyrAspHisGlyThrAlaH 786
2194 TAACTGACTTGGACAGCTCTCGGGATGATATGACCATGGACAGCTC 2243
786 isLysTyrIleIleArgIleSerThrSerIleLeuAspLeuArgAspLys 802
2244 ACAAGTATATCATTCGAATAAGTAACTATTTCTGATCTCAGACACAAG 2293
803 PheAsnGluSerLeuGlnValAsnThrThrAlaLeuIleProLysGluAl 819
2294 TTCAATGAATCTCTCAAGTGAATCTACTGCTCTCATCCCAAGGAAGC 2343
819 aAsnSerGluGluValPheLeuPheLysProGluAsnIleThrPheGluA 836
2344 CAACTCTCAGGAAGTCTTTTGTGTTTAAACACAGAAACATTAATTTGAA 2393
836 sNGLyThrAspLeuPheIleAlaIleGlnAlaValAspLysValAspLeu 852
2394 ATGGCACAGATCTTTTCAATGCTATTCAGGCTGTTGATAAGGTCGATCTG 2443
853 LysSerGluIleSerAsnIleAlaArgValSerLeuPheIleProGluI 869
2444 AAATCAGAAATATCCAACTTGCAGAGTATCTTTGTTTATTCCTCCACA 2493
869 nThrProGluThrProSerProAspGluThrSerAlaProCysProA 886
2494 GACTCCGCCAGACACACCTAGTCTGTGATGAAGAGTCTGCTCTTGCTCTA 2543
886 sNleHisIleAsnSerThrIleProGlyIleHisIleLeuLysIleMet 902
2544 ATATTTCATCAACAGCAGCACTTCTGTCATTCATTCATTTTAAAAATTATG 2593
903 TrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914
2594 TGGAAAGTGGATAGGAACTGACGCTGTCATATAGCC 2629
seq_name: /SDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH46101
seq_documentation_block:
ID AAH46101 standard; DNA: 2739 BP.
XX
AC AAH46101:
XX
XX
DT 11-SEP-2001 (first entry)
XX
DE Mouse Gob-5 coding sequence, SEQ ID NO:3.
```

```
XX Mouse Gob-5; murine; goblet cell; human CICAL orthologue; drug screening;
KW expression inhibition; antisense therapy; gene therapy; bronchial asthma;
KW chronic obstructive pulmonary disease; antiasthmatic; ds.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 1..2739
FT FT /*tag= a
FT FT /partial
FT FT /product= "Mouse Gob-5"
FT FT /note= "No stop codon given in the specification"
XX
PN WO200138530-A1.
XX
XX 31 MAY 2001.
XX
XX 22 NOV 2000; 2000WO-JP08232.
XX
XX 24 NOV 1999; 99JP-0333479.
XX
XX 27 APR 2000; 2000JP-0127589.
XX
XX (TAKE ) TAKEDA CHEM IND LTD.
XX
XX Nakanishi A, Morita S;
XX
XX WPI; 2001-355935/37.
XX
XX P-IPDB; AAB73715.
XX
XX New antisense nucleotide, useful for treatment and prevention of
XX bronchial asthma and chronic obstructive pulmonary disease -
XX
XX Claim 3; Page 80-82; 104pp; Japanese.
XX
XX The invention relates to an antisense nucleotide targetted to the mouse
XX Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
XX the CICAL gene (coding sequence shown in AAH46102). The invention also
XX relates to an antibody specific for the Gob-5 protein, medical and
XX diagnostic compositions containing the antisense nucleotide or the
XX antibody, and methods and kits for screening for compounds which inhibit
XX the protein. Gob-5 and CICAL are proteins expressed by goblet cells.
XX The antisense oligonucleotides and antibody are therefore useful for the
XX treatment and prevention of bronchial asthma and chronic obstructive
XX pulmonary disease. The present sequence represents the mouse Gob-5
XX gene coding sequence.
XX
SQ Sequence 2739 BP; 784 A; 687 C; 655 G; 613 T; 0 other;
```

```
alignment_scores:
Quality: 3656.50 Length: 915
Ratio: 4.337 Caps: 4
Percent Similarity: 92.131 Percent Identity: 75.847
alignment_block:
US-09-049-696-41 x AAH46101
Align'seg 1/1 to: AAH46101 from: 1 to: 2739
1 MetGlyPropheLysSerSerValPheIleLeuIleLeuHisLeuLeuG1 17
1 ATGGAATCTTTGAGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGGA 50
17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34
51 AGGAGTCTGAGTCCCTCATCTCACTGAACCAACAGCGCTAAGG 100
34 lyIleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
101 GCATCGTCATCGCCATAGACACGACGTCGCCGGAAGATGAAGCCCTCAT 150
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
```

151 CAACACATAAAGGACATGGTGTGATCAGGCGCTCTCCATACCTGTTTGAAGC 200
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluT 84
201 TACAGGAAAAAGATTTTACTTCAAAATGTGCCATTTTGTATCCCGAGA 250
84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
251 GCTGGAAGCAAGCCGTGAATACAGAGCCAAAACCTTGAACCTTCAAA 300
101 AsnAlaAspValLeuValAlaGluSerThrProProGlyAsnAspGluPr 117
301 AACGCTCATCTCTTATCAACAACAGCCCTCTAGGCAATGATGAGCC 350
117 oTyrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134
351 CTACACCGAACATATAGGACATGTGGAGAAAAGGGATCAGGATTCACC 400
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
401 TGACTCTGACTCTTAGCAGGAAAGAGCTGACTCAGTATGGGCCACAA 450
151 GlyArgAlaPheValHisGluTyrAlaHisLeuArgTrpGlyValPheAs 167
451 GACAGGACCTTTGTCCATGAGTGGGCTCACTCCGATGGGGAGTGTATA 500
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGlnA 184
501 TGAATACAAACAGCAGCAGAGTTCATCTATCCAAAGGAAAACCCCAAG 550
184 laValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLysCys 200
551 CAGTGAGGTCTCAGCAGCCATACCGGTAAATCAAGTTTCGTCGTGC 600
201 GlnGlyGlySerCysTyrThrLys...ArgCysThrPheAsnLysValTh 216
601 CAGGAGGACAGTGTATCATCTACGGAAGTGTGTAATCGACAGAGTAAC 650
216 rGlyLeuTyrGlyLysGlyCysGluPheValLeuGlnSerArgGlnThrG 233
651 GGGACTATATAAGACAATTTGTATTGTACCAGATCCACACCAAAACG 700
233 lLysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPhe 249
701 AGAAGGCTCCATCATGTTTAAACCAAAATATCAATCTGTGGTGAATTC 750
250 CysThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLy 266
751 TGTACAGAAAAAATCAATCAAGAAGCCCAAAATGACCAAAACCAACG 800
266 sCysAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheL 283
801 ATGCAATCTCCGAAGCAGTGGGAAGTCATCCAGGAATCTGAGGACTCA 850
283 ysLysThrThrProMetThrThrGlnProProAsnProThrPheSerLeu 299
851 AGCAAAACCATCTCCATGACAGCCAGCCACCTGCACCCACCTTCTCAGT 900
300 LeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySe 316
901 CTGCAATTTGGACAAAGAAATGTGTGCTTAGTCTTGTATAGTCCGGAG 950
316 rMetAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuP 333
951 CATGCTGAACCATGATCGTCTTAAACCAATGAATCAGGCAAGCCGGCTTT 1000
333 heLeuLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPhe 349
1001 TCCGTGCTCAGACATGTGGAGCAGGAGCCCTGGGGTGGGATGGTACCTTT 1050
350 AspSerAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySe 366

1051 GACACTGCTGCTATGTACAAAGCGAACTCAACAGTTTAAACAGTGGTGCC 1100
366 rAspArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyT 383
1101 TGACAGAGATCTGTGATCAAGCACTTTACCACAGTATCTCGAGGAGGA 1150
383 hrSerIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLys 399
1151 CATCTATATGCTCTGGCCTTCGGACAGCATTTACAGTGTATAAAGAAAG 1200
405 TyrProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAs 416
1201 TATCCAACATGATGGATCTGAAATGCTGTGTCGACCGATGGGAGGACAA 1250
416 nThrIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleH 433
1251 CACCATTAGCAGCTGCTTTGACCTGGTGAAGCAGAGCGGGCCATCATCC 1300
433 isThrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluLeuSer 449
1301 ATACAGTGGCCCTGGAGCGGCTGCGCTAAAGAGCTTGAGCAGCTGTCC 1350
450 LysMetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnAs 466
1351 AAATGACAGGAGGCTGCAGACATACCTCTCGGATCAGTTTCAGAACAA 1400
466 nGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValS 483
1401 TGGTCTTGTGTATGCTTTGCGACACCTCTCTCAGGAAATGCGCGCATCG 1450
483 exGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSer 499
1451 CTCAGACCTCCATCCAGCTGGAGAGCGGGAGTTAATCTCCAGAAATAC 1500
500 GlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspTh 516
1501 CAATGATCAATGGCTCAGTGTGTCGACAGCTCGTGGCAAGACAC 1550
516 rLeuPheLeuIleThrTrpThrThrGlnProGlnIleLeuLeuTrpA 533
1551 CTGTGTTCTTATCAGCTGACACAGCTCTCTCTCAATATTTATCTGGG 1600
533 sProSerGlyGlnLysGlnGlyPheValValAspLysAsnThrLys 549
1601 ATCCAGCGGAGTGGAAACAAATGTTTATATAGTACACACACCACTAAG 1650
550 MetAlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLysTy 566
1651 GTGGCTTACCTCCAAGTCCAGCAGCGCTAAGTTGGCTTTTGGAAATA 1700
566 rSerLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSerArgA 583
1701 CAGCATTTCAAGCAGCTCAGACTCTCAGCTTGACTGTACCTCCCGTG 1750
583 laSerAsnAlaThrLeuProProIleThrValThrSerLysThrAsnLys 599
1751 CAGCAAGTCTACACTGCTCTCTATTACAGTGACCCCGGTAGTGAATAAG 1800
600 AspThrSerLysPheProSerProLeuValValTyrAlaAsnIleArgG 616
1801 AACACAGGAAATTCGCCAGCCCTGTACAGTGTATGCAAGCATTCGCCA 1850
616 nGlyAlaSerProIleuArgAlaSerValThrAlaLeuIleGluSerV 633
1851 AGAGGCTCGCCTATTCTCAGGCGCCAGCTCAGCCCTTGATTGAATCTG 1900
633 alAsnGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAla 649
1901 TGAATGGAHAACAGTAACCTGGAATTAAGTAAACGAGCAGGTGCC 1950
650 AspAlaThrLysAspGlyValTyrSerArgTyrPheThrThrTyrAs 666
1951 GATGCCACCAAGAATGATGCTGTCTACTCAAGGTTTTTTACAGCTTTTGA 2000

```

666 pThrAsnGlyArgTyrSerValLysValArgAlaLeuGlyGlyValAsnA 693
|:::|||||||:::|||||||:::|||||||:::|||||||:::
2001 TCCAATGGTAGATACAGCGTTAAAATATGGGCTCTGGGAGGAGTCACCT 2050
|:::|||||||:::|||||||:::|||||||:::|||||||:::
683 IaAlaArgArgVAlIleProGlnGlnSerGlyAlaLeuTyrllePro 699
|:::|||||||:::|||||||:::|||||||:::|||||||:::
2051 CAGACAGACAGAGACGCACCTCCGAAGAAGACAGACCATTGTACATAGAT 2100
|:::|||||||:::|||||||:::|||||||:::|||||||:::
700 GlyTrpIleGluAsnAspGluleGlnTrpAsnProProArqProGluIl 716
|:::|||||||:::|||||||:::|||||||:::|||||||:::
2101 GGCTGGATTGAGGATGGTCAAGTAAGAAATGAACCCACCACGTCCTCAAAC 2150
|:::|||||||:::|||||||:::|||||||:::|||||||:::
716 eAsnLysAspValGlnHisLysClnValCysPheSerArgThrSeriS 733
|:::|||||||:::|||||||:::|||||||:::|||||||:::
2151 TAGTTAT.....GTTCAAGACAAGCAGGTGCTTCAGCAGGACATCTT 2194
|:::|||||||:::|||||||:::|||||||:::|||||||:::
733 erGlyGlySerPheValAlaSerAspValPro...AsnAlaProIlePro 748
|:::|||||||:::|||||||:::|||||||:::|||||||:::
2195 CAGGGGGATCGTTTGTCGCCACCAATGTCCCAGCACAGCTCCCATTCCT 2244
|:::|||||||:::|||||||:::|||||||:::|||||||:::
749 AspLeuPheProProGlyGlnIleThrAspLeuLysAlaGluIleHisGl 765
|:::|||||||:::|||||||:::|||||||:::|||||||:::
2245 GACCTCTTTCCACCCCTGTCAAAATCACTGACCTGAAGCCAGCAGTCCAAGG 2294
|:::|||||||:::|||||||:::|||||||:::|||||||:::
765 yGlySerLeulleAsnLeuThrTrpThrAlaProGlyAspAspTyrlle 782
|:::|||||||:::|||||||:::|||||||:::|||||||:::
2295 GCAGAACCCTGGTGAATCTACGTGGACGGCTCTCGGGATGACTACGACC 2344
|:::|||||||:::|||||||:::|||||||:::|||||||:::
782 IsGlyThrAlaHisLysTyrlleIleArylleSerThrSeriLeuLeuAsp 798
|:::|||||||:::|||||||:::|||||||:::|||||||:::
2345 ACGGGAGAGCTCCCAACTACATCCGAATGAGCAGCAGTATCGTTGAT 2394
|:::|||||||:::|||||||:::|||||||:::|||||||:::
799 LeuArgAspLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuIl 815
|:::|||||||:::|||||||:::|||||||:::|||||||:::
2395 CTCAGGGACCACTTCAACACCTCACTCCAAGTGAACACTACCGGTCCTAT 2444
|:::|||||||:::|||||||:::|||||||:::|||||||:::
815 eProLysGluAlaAsnSerGluGluValPheLeuPheLysProGluAsnI 832
|:::|||||||:::|||||||:::|||||||:::|||||||:::
2445 CCCCAAGAGGCCAGCTCTGAGGAATCTTGNAGTTTGAACCTGGGAGGCA 2494
|:::|||||||:::|||||||:::|||||||:::|||||||:::
832 leThrPheGluAsnGlyThrAspLeupheIleAlalleGlnAlaValAsp 848
|:::|||||||:::|||||||:::|||||||:::|||||||:::
2495 ACACCTTTTGAAATGGCACAGATATCTTCATTGCTATCCAGGCTGTGGAT 2544
|:::|||||||:::|||||||:::|||||||:::|||||||:::
849 LysValAspLeuLysSerGluIleSerAsnIleAlaArgValSerLeuPh 865
|:::|||||||:::|||||||:::|||||||:::|||||||:::
2545 AAGTCCAAATCTCAAAATCAGAAATCTCCAACATTCACGGGTGCTGTGTT 2594
|:::|||||||:::|||||||:::|||||||:::|||||||:::
865 eIleProGlnThrProGluThrProSerProAspGluThrSeriA 882
|:::|||||||:::|||||||:::|||||||:::|||||||:::
2595 CATCCCCGCTCAGGAGCGG.....CCCATTCCCGAGAGCTCAACTC 2635
|:::|||||||:::|||||||:::|||||||:::|||||||:::
882 laProCysProAsnIleHisIleAsnSerThrIleProGlyIleHisIle 898
|:::|||||||:::|||||||:::|||||||:::|||||||:::
2636 CCCCTTGCTCAGCATCAGCATCAACAGCAGCACCATTCTCGCATCCACGTG 2685
|:::|||||||:::|||||||:::|||||||:::|||||||:::
899 LeuLysIleMetTrpLysTrpIleGlyGluLeuGlnLeuSeriIle 913
|:::|||||||:::|||||||:::|||||||:::|||||||:::
2686 CTGAAGATAAATGTGAAGHGGGTAGGGGAATCCAGCATGCACACTA 2730
|:::|||||||:::|||||||:::|||||||:::|||||||:::
seq_name: /SID$2/qcgdata/geneseq/geneseqn/NA2001.DAT :AAH461120

```

seq_documentation_block:

ID	AAH46120 standard	seq_documentation_block	CDNA	2843 BP
1	AAH46120 standard	seq_documentation_block	CDNA	2843 BP

ID	AAH46120
XX	
AC	AAH46120;

DT 11-SEP-2001 (first entry)

DE Mouse Gob-5 cDNA, SEQ ID NO:22.

DE	Mouse Gob-5 CDNA, SEQ ID NO:22.
XX	
KW	Mouse Gob-5; murine; goblet cell; human CLCA1 orthologue; drug screening;

KW	expression inhibition; gene therapy; bronchial asthma;
KW	chronic obstructive pulmonary disease; antiasthmatic; ss.
XX	
XX	Mus sp.
OS	
XX	
FH	Location/Qualifiers
FT	15..2756
FT	CDS
FT	/tag= a
FT	/product= "Mouse Gob-5"
XX	
XX	WO200138530-A1.
PN	
PD	31 MAY-2001.
XX	
XX	22 NOV-2000; 2000WO-JP08232.
PF	
XX	
XX	24 NOV-1999; 99JP-0333479.
PR	
PR	27 APR-2000; 2000JP-0127589.
XX	
XX	(TAKE) TAKEDA CHEM IND LTD.
PA	
XX	
XX	Nakanishi A., Morita S;
PI	
XX	
DR	WPI: 2001-355935/37.
DR	P-PSTDB; AAB73715.
XX	
XX	New antisense nucleotide, useful for treatment and prevention of
FT	bronchial asthma and chronic obstructive pulmonary disease -
PT	
XX	
PS	Example 1; Page 89-91; 104pp; Japanese.
XX	
CC	The invention relates to an antisense nucleotide targetted to the mouse
CC	Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
CC	the CLCA1 gene (coding sequence shown in AAH46102). The invention also
CC	relates to an antibody specific for the Gob-5 protein, medical and
CC	diagnostic compositions containing the antisense nucleotide or the
CC	antibody, and methods and kits for screening for compounds which inhibit
CC	the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.
CC	The antisense oligonucleotides and antibody are therefore useful for the
CC	treatment and prevention of bronchial asthma and chronic obstructive
CC	pulmonary disease. The present sequence represents mouse Gob-5 cDNA.
XX	
SQ	Sequence 2843 BP; 815 A; 705 C; 680 G; 643 T; 0 other;

alignment_scores:

Alignment Scores:			
Quality:	3656.50	Length:	915
Ratio:	4.337	Gaps:	4
Percent Similarity:	92.131	Percent Identity:	75.847

alignment block:

US-09-049-696-41 x AAH46120

Align: seq 1/1 to: AAH46120 from: 1 to: 2843

1 MetGlyProPheLysSerSerValPheIleLeuIleLeuHisLeuLeuG1
 115 ATGGAAATCTTTGAAGAGTCCTGCTCTCTCTCTGTATCTCCACCTTCGGA 64
 17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34
 65 AGGAGTCTCTGAGTCAGTCCCTCATCCAACTCAACAACAACGGCTATCAGG 114
 34 lyIleValValAlaIleAspProAsnValProGluAspGluThrLeuLe 50
 115 GCATCGTCATCGCCCATAGACCACGACGTGCGGGAAGATGAAGCCCTCATTT 164
 51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
 165 CAACACATAAAGGACATGGTGACTCAGCGCCTCCATACCTGTTTGAAGC 214
 67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluT 84

215 TACAGGAAAAAGATTCTTCAAAAATGTTGCCATTTTATTCCCGAGA 264
84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
265 GCTGGAAGCAAGCCGTGAATATACAGAGCCAAACCTTGAAACCTTCAA 314
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
315 AAGCCTGATGCTCTGTATCAACACAGCCCTCTAGGCAATGATAGCC 364
117 ofYrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134
365 CTACACCGAACATATAGGAGCATGTGGAGAAAAGCGCATCAGGATTCACC 414
134 euThrProAspPheIleAlaGlyLysLeuAlaGluTyrGlyProGln 150
415 TGACTCCTGACTTCTTAGCAGGAAAGCTGACTCAGTATGGCCACAA 464
151 GlyArgAlaPheValHisGluTyrPheAlaHisLeuArgTyrPheValPheAs 167
465 GACAGACCTTGTCCATGATGGGCTCACTTCCGATGGGAGTGTTAA 514
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGlnA 184
515 TGAATACAACACAGCAGAGAAAGTTCTACTTATCCAAAGAAAACCCCAAG 564
184 laValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLysCys 200
565 CAGTGAAGGTTCAGCAGCCATTTACCGGTAAATAATCAAGTTCGCGTGC 614
201 GlnGlyGlySerCysTyrThrLys...ArgCysThrPheAsnLysValTh 216
615 CAGGAGGAGCAGTTGATCACTACGGAAGAGTGTGTAATCGACAGAGTAAC 664
216 rGlyLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrG 233
665 GGCAGCTGTATAAGACAAATGTGTATTGTACCAGATCCACACCAAAACG 714
233 luLysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPhe 249
715 AGAAGGCTTCCATCATGTTTAAACCAAAATATCAATTTCTGTGGTTGAATC 764
250 CysThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLy 266
765 TGTACAGAAAAAATCACAATCAAGAAGCCCAAAATGACCAAAACCAAG 814
266 sCysAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheL 283
815 ATGCAATCTCCGAAGCACGTGGGAAGTCATCCAGGAATCTGAGGACTTCA 864
283 ysLysThrThrProMetThrThrGlnProProAsnProThrPheSerLeu 299
865 AGCAAAACCACTCCATGACAGCCGCCACCTGACCCACCTTCTCACGTG 914
300 LeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySe 316
915 CTGCAAAATGGACAAAAGATTGTGCTTAGTCTTGTGATAAGTCCGGAG 964
316 rMetAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuP 333
965 CATGCTCAACGATGATCGCTTAAACGAATGAATCAGGAAGCGCGCTTT 1014
333 heLeuLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPhe 349
1015 TCCTGTGTCAGACTGTGGAGCAGGGATCTGGGTGGGATGTGACCTTT 1064
350 AspSerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySe 366
1065 GACAGTGTGCTCTATCTACAAAGCAAGTCAAAACAGTTAAACAGTGGTGC 1114
366 rAspArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyT 383
1115 TGACAGAGATCTGCTGATCAAGCACTTACCACAGTATCTGCGAGGAGGA 1164

383 hrSerIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLys 399
1165 CATCTATATGCTCTGGCCCTCGACAGCATTACAGTGTATAAAGAAGAAG 1214
400 TyrProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAs 416
1215 TATCCAACTGATGGATCTGAATTTGCTGCTGACCGATGGGAGGACAA 1264
416 nThrIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleH 433
1265 CACCATTAGCAGCTGCTTTGACCTGCTGAAGCAGAGCGGGCCATCATCC 1314
433 isThrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluLeuSer 449
1315 ATACAGTGGCCCTGGAGCCGCTGCGCTAAAGAGCTTGAGCAGCTGTCC 1364
450 LysMetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnAs 466
1365 AAAATGACAGGAGCCCTGCAGACATACTCTTCGATCAGGTTTCAGAACAA 1414
466 nGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValS 483
1415 TGTCTCTGTTGATGCTTTCGACGACTCTCTCAGCAATGCGGCGATCG 1464
483 erClnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSer 499
1465 CTCAGCAGCTCCATCCAGCTGGAGAGCAGGGAGTTAATCTCCAGAATAAC 1514
500 GlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspTh 516
1515 CAATGATGAATGGCTCAGTATGCTGGACAGCTCGTGGCAAGGACAC 1564
516 rLeuPheLeuIleThrTrpThrGlnProGlnIleLeuLeuTrpA 533
1565 CTGTGTTCTTATCACCTGGACACAGCATCTCTCTACAATATTTATCTGG 1614
533 spProSerGlyGlnLysGlnGlyPheValValAspLysAsnThrLys 549
1615 ATCCAGCGGAGTGGAAACAAATGTTTATCTAGACACACACCACTAAG 1664
550 MetAlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLysTy 566
1665 GTGGCTACCTCCAGTCCAGCAGCGCTAGGTTGGCTTTTGGAAATA 1714
566 rSerLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSerArgA 583
1715 CAGCATTCAGGAGGAGCTCAGAGACTCTCAGCTGACTGTCACTCCCGTG 1764
583 laSerAsnAlaThrLeuProIleThrValThrSerLysThrAsnLys 599
1765 CAGCAAGTGTACACTGCTCTCTATTACAGTGACCCCGGTAGTGAATAAG 1814
600 AspThrSerLysPheProSerProLeuValValTyrAlaAsnIleArgG 616
1815 AACACAGGAAATTCGCCAGCCCTGTACAGCTGTATGCAAGCATTCGCCA 1864
615 nGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerV 633
1865 AGGAGCTCGCCTATTCTCAGGCCAGCGCTCACAGCTTGTGATTGAATCTG 1914
633 aLysGlyLysThrValThrLeuLeuLeuAspAsnGlyAlaGlyAla 649
1915 TGAATGGAAAAACAGTAACTTGGAAATTTACTGGATACCGAGCAGGTGCC 1964
650 AspAlaThrLysAspGlyValTyrSerArgTyrPheThrThrTyrAs 666
1965 GATGCCCAACCAAGATGCTGCTACTCAAGGTTTTTTACAGCTTTTGA 2014
666 pThrAsnGlyArgTyrSerValLysValArgAlaLeuGlyGlyValAsnA 583
2015 TGCAAAATGATAGACAGCTTAAAAATATGGGCTCTGTGGAGGAGTCACTT 2064

```

683 laAlaArgArgValIleProGlnGlnSerGlyAlaLeuTyrIlePro 699
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
2055 CAGCAGACAGAGACACCTCGAAGAAGACAGACCATCATAGAT 2114
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
700 GlyTrpIleGluAsnAspGluIleGlnTrpAsnProProArgProGluIle 716
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
2115 GCCTGGATTGAGGATGGTGAAGTGAAGATGAACCCACCCACCTCTGGAAC 2164
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
716 eAsnLysAspAspValGlnHisLysGlnValCysPheSerArgThrSers 733
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
2165 TAGTTAT.....GTTCAAGACAAGACAGCTGTCTTCAGCAGGACATCTT 2208
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
733 erGlyGlySerPheValAlaSerAspValPro...AsnAlaProIlePro 748
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
2209 CAGGGGATCTTGTGTGCCCAACATGTCCTCCGAGCAGCTCCCATCTCT 2258
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
749 AspLeuPheProProGlyGlnIleThrAspLeuLysAlaGluIleHisG1 765
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
2259 GACCTCTTTCACCTGTCAAATCAGCTGACCTGAAGGCCACCATCCAAAGG 2308
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
765 yGlySerLeuIleAsnLeuThrTrpAlaProGlyAspAspTyrAspH 782
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
2309 GCAGAACCTGTGTAATCTGACGTGACGGCTCTCTGGGATGACTACGACC 2358
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
782 isGlyThrAlaHisLysTyrIleIleArgIleSerThrSerIleAsp 798
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
2359 ACGGAGAGCTTCAACTACATACATCCGAATGAGCACCAGATCGTTGAT 2408
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
799 LeuArgAspLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuI1 815
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
2409 CTCAGGGACCACTTCAACACCTCACTCCAACTGAACACTACCGGTCTAT 2458
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
815 eProLysGluAlaAsnSerGluGluValPheLeuPheLysProGluAsnI 832
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
2459 CCCAAAGAGAGCCAGCTCTGAGGAATCTTTGAGTTTGAAGCTGGGAGGCA 2508
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
832 leThrPheGluAspGlyThrAspLeuPheIleAlaIleGlnAlaValAsp 848
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
2509 ACATTTTGGAAATGGCAGATATCTTCATGTCATCCAGGCTGTGAT 2558
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
849 LysValAspLeuLysSerGluIleSerAsnIleAlaArgValSerLeuPh 865
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
2559 AAGTCCAATCTGAATCAGAAATCTCCAACATTCACAGGGTGTCTGTCT 2608
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
865 eIleProGlnThrProProGluThrProSerProAspGluThrSerA 882
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
2609 CATCCCGCTCAGGAGCGG.....CCCATTCGCCGAGAGACTCACTC 2649
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
882 laProCysProAsnIleHisIleAsnSerThrIleProGlyIleHisIle 898
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
2650 CCCCTTGTCTGACATCAGCATCAGCACCACCATCTCTGGCATCCAGTG 2699
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
899 LeuLysIleMetTrpLysTrpIleGlyGluLeuGlnLeuSerIle 913
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
2700 CTGAAGATAATCTGGAAGTGGCTAGGGGAAATGCAAGGTGACACTA 2744
  ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||

```

seq_name: /SDS2/gcdata/geneseq/geneseqn/NA1999.DAT:AAF81925

seq_documentation_block:

AAF81925 standard; cDNA; 2931 BP.

AAF81925;

13-JUN-2001 (first entry)

Murine ICACC-1 nucleotide sequence.

ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
interleukin 9 induced calcium activated chloride channel; IL-9;
calcium activated chloride channel; anti-allergic; anti-asthmatic;
anti-inflammatory; immunomodulatory; cystic fibrosis;
inflammatory bowel disease; autoimmune disease; ss.

OS Mus sp.
FH Réy
FT CDS
FT Location/Qualifiers
FT 8..2749
FT /*tag= a
FT /product= "ICACC-1"
FT /note= "IL-9 induced calcium activated chloride channel"

W09944620-A1.
XX
PD 10-SEP-1999.

XX 03-MAR-1999; 99WO-US04703.

XX 03-MAR-1998; 98US-0076815.

XX (MAGA-) MAGAININ PHARM INC.

XX Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;
PI Nicolaidis NC, Zhou Y, Dong Q;

XX WEI; 1999-550979/46.

DR P-PSDB; AAB74822.

XX New nucleic acid encoding calcium activated chloride channel, used to
PT identify, e.g. specific modulators for treating atopic allergy -

XX Claim 2; Fig 2; 75pp; English.

XX The present sequence encodes the murine interleukin 9 (IL-9) induced
CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
CC have anti-allergic, anti-asthmatic, anti-inflammatory and
CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
CC used to alleviate asthma (or more generally atopic allergy), while those
CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
CC (increased levels) or IBD (reduced levels), also for monitoring
CC treatment of these conditions. The ICACC proteins can be used:
CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
CC reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
CC identify modulators and binding partners. ICACC polynucleotides can be
CC used to generate transgenic animals or recombinant cells, used to screen
CC for antagonists, also as a source of therapeutic antisense agents or
CC diagnostic probes (for quantifying mRNA expression, e.g. for
CC identification of modulators).

XX Sequence 2931 BP; 867 A; 719 C; 684 G; 661 T; 0 other;

alignment_scores:
Quality: 3656.50 Length: 915
Ratio: 4.337 Gaps: 4
Percent Similarity: 92.131 Percent Identity: 75.847

alignment_block:
US-09-049-696-41 x AAF81925

Align seg 1/1 to: AAF81925 from: 1 to: 2931

1 MetGlyProPheLysSerSerValPheIleLeuIleLeuHisLeuGlu 17
||| ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
3 ATGGAATCTTTGAAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGGA 57
17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34
||| ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
58 AGGAGTCTGAGTGAGTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGAGG 107
34 lyIleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
||| ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||
108 GCATCGTCATCGCATAGACACCCAGCAGCTGCGGAGAGATGAAGCCCTCAT 157
||| ::::::::::: ||| ::::::::::: ||| ::::::::::: ||| ::::::::::: |||

134 euThrProAspPheIleAlaGlyLysLeuAlaGluTyrGlyProGln 150
 :|||:::|||||:::|||||:::|||||:::|||||:::|||||
 422 TCACCCCTGCAGCTTCCTACTTGGAATAAAATGAATATGGACCACCA 471

151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
 |||:::|||||:::|||||:::|||||:::|||||:::|||||
 472 GCGAACCTGTGTGCCATAGTGGGCTCACCTCGGTGGGAGTGTTTGA 521

167 pGluTyrAsnAsnAspGluLysPheThrLeuSerAsnGly...ArgIle 183
 :|||:::|||||:::|||||:::|||||:::|||||:::|||||
 522 TGAGTACAATGAAGATCAGCCTTCTACCGTGCTTAAGTCAAAAAAATCG 571

183 InAlaValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLys 199
 ::||:::|||||:::|||||:::|||||:::|||||:::|||||
 572 AAGCAACAAGGTGTCCGACGATCTCTGGTAGAAAATAGATTATAAG 621

200 CysGlnGlnGlySerCysTyrThrLysArgCysThrPheAsnLysValTh 216
 |||:::|||||:::|||||:::|||||:::|||||:::|||||
 622 TGTCAAGGAGCAGCGTCTTAGTAGAGCATGCGAGAATTGATTCTACAAC 671

216 rGlyLeuTyrGluLysGlyCysGluPheValLeuGlnInserArgGlnThrG 233
 |||:::|||||:::|||||:::|||||:::|||||:::|||||
 672 AAAACTGTATGGAAAAGATTGTCAATTCTTTCCTGTATAAGTACAACAG 721

233 IuLysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPhe 249
 |||:::|||||:::|||||:::|||||:::|||||:::|||||
 722 AAAAGCATCCATAATGTTTTATGCAAAAGTATTGATTCTGTGTGAATTT 771

250 CysThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLy 266
 |||:::|||||:::|||||:::|||||:::|||||:::|||||
 772 TGTACAGAAAAAACCCATAATCAAGAAAGCTCCAAGCCTCAAAAACATAAA 821

266 scYasNLeuArgSerThrTpGluValIleArgAspSerGluAspPheL 283
 |||:::|||||:::|||||:::|||||:::|||||:::|||||
 822 GTGCAATTTTTAGAAGTACATGGGAGGTGATTAGCAATTCGAGGATTTTA 871

283 ySLysThrThrProMetThrThrGlnProProAsnProThrPheSerLeu 299
 |||:::|||||:::|||||:::|||||:::|||||:::|||||
 872 AAACACCATACCCATGGTGACACCACCTCCTCCACCTGTCTTCTCATTTG 921

300 LeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySe 316
 |||:::|||||:::|||||:::|||||:::|||||:::|||||
 922 CTGAAGATCAGTCAAAGAATTTGTGTGTTAGTTCTTGATAGTCTCGAAG 971

316 rMetAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyClnLeup 333
 |||:::|||||:::|||||:::|||||:::|||||:::|||||
 972 CATGGGGGGTAAAGCCGCCTTAAATCGAATGAAATCAAGCAGCAAAACATT 1021

333 heLeuLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPhe 349
 |||:::|||||:::|||||:::|||||:::|||||:::|||||
 1022 TCCTGCTGCACACTGTTGAAAATGGATCCTGGGTGGGATGTTTCACTTT 1071

350 AspSerAlaAlaHisValGlnSerGluLeuIleGlnLeuAsnSerGlySe 366
 |||:::|||||:::|||||:::|||||:::|||||:::|||||
 1072 GATAGTACTGCCACTATTGTATAAAGCTAATCAAAATAAAAAGCAGCTGA 1121

366 rAspArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyClyt 383
 :|||:::|||||:::|||||:::|||||:::|||||:::|||||
 1122 TGAAGAAACACACTCATTCGGCAGGATTACCTTACATTCCTCTGGGAGGAA 1171

383 hrSerIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLys 399
 |||:::|||||:::|||||:::|||||:::|||||:::|||||
 1172 CTTCCATCTGCTCGGAATTAATAATATGCATTTACAGTTGTTGGAGCTA 1221

400 Tyr...ProfThrAspGlySerGluIleValLeuLeuThrAspGlyGluAs 415
 :|||:::|||||:::|||||:::|||||:::|||||:::|||||
 1222 CATTTCCCAACTCGATGGATCCGAAGTACTGCTGCTGACTGATGGGAGGA 1271

415 pAsnThrIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleI 432
 |||:::|||||:::|||||:::|||||:::|||||:::|||||
 272 TAACACTGCAAGTTCTGTATTTGATGAAGTGAACAAGTGGGCGCAATTG 1321

[illegible]

```
|||||.....|||||.....|||||.....|||||.....|||||.....
2219 GAACAGATCCGGAGTGATTTGGTATCACAGTCCCAAGCCTTCC 2268
747 ILeProAspLeuPheProGlyGlnIleThrAspLeuAlaGlu1 763
:::|||||:::|||||:::|||||:::|||||:::|||||:::
2269 TTGCTGACCAATACCCACCAAGTCAATCACAGACCTTGATGCCACGT 2318
763 eHisGlyGlySerLeuIleAsnLeuThrPrpAlaProGlyAspAsPT 780
:::|||||:::|||||:::|||||:::|||||:::|||||:::
2319 TCATGAGGATAAGATTATT...CTTACATGGACAGCACCAGGAGATAATT 2365
780 yRAspHisGlyThrAlaHisLysTyrlleIleArgIleSerThrSerIle 796
:::|||||:::|||||:::|||||:::|||||:::|||||:::
2366 TTGATGTGGAAGTTCAACGTTATATCATAGATAAGTCAAGTATT 2415
797 LeuAspLeuArgAspLysPheAsnGluSerLeuGlnValAsnThrAl 813
|||||.....|||||.....|||||.....|||||.....|||||.....
2416 CTTGATCTAAGACAGACAGTTTGTATGATGCTCTCTCAAGTAATACTACTGA 2465
813 aLeuIleProLysGluAlaAsnSerGluGluValPheLeuPheLysProG 830
|||||.....|||||.....|||||.....|||||.....|||||.....
2466 TCTGTACCAAGAGAGGCCAATCTCCAGGAAAGCTTTGCATTTAACCAG 2515
830 luAsnIleThrPheGluAsnGlyThrAspLeuPheIleAlaIleGlnAla 846
|||||:::|||||:::|||||:::|||||:::|||||:::
2516 AAATATCTCAGAAGAAATGCAACCCACATATTATTGCGCATTAAGT 2565
847 ValAspLysValAspLeuLysSerGluIleSerAsnIleAlaArgValse 863
:::|||||:::|||||:::|||||:::|||||:::|||||:::
2566 ATAGATAAAGCAATTTGACATCAAAAGTATCAACATTCGCAAGGVAAC 2615
863 rLeuPheIleProGlnThrProGlu...ThrProSerProAspG 879
|||||.....|||||.....|||||:::|||||:::
2616 TTTGTTATCTCTCAAGCAAAATCTGATGACATTGATCTACACCTACTC 2665
879 luThrSerAlaProCysProAsnIleHisIleAsnSerThrIleProGly 895
|||:::|||||:::|||||:::|||||:::
2666 CTACTCTACTCTCTCTCTGATAAAGTCAATAATCT.....GGA 2706
896 IleHisIleLeuLysIleMetTrpLysTrpIleGlyGluLeu 909
:::|||||:::|||||:::|||||:::
2707 GTTAATATTTCTACGCTGGATTCTCTGATGGTCTGTT 2748
seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAF92092
seq_documentation_block:
ID_AAF92092 standard; cdna; 3265 BP.
XX
AC AAF92092;
XX
15-MAY-2001 (first entry)
XX
Human PRO1124 cdna.
XX
Human; PRO protein; mapping; ss.
XX
Homo sapiens.
XX
WO200116318-A2.
XX
08-MAR-2001.
XX
24-AUG-2000; 2000WO-US23328.
XX
01-SEP-1999; 99WO-US20111.
XX
15-SEP-1999; 99WO-US21090.
XX
07-DEC-1999; 99US-0169495.
XX
09-DEC-1999; 99US-0170262.
XX
11-JAN-2000; 2000US-0175481.
XX
18-FEB-2000; 2000WO-US04341.
XX
18-FEB-2000; 2000WO-US04342.
XX
22-FEB-2000; 2000WO-US0414.
XX
01-MAR-2000; 2000WO-US05601.
```

```
PR 03-MAR-2000; 2000US-0187202.
PR 25-APR-2000; 2000US-0199397.
PR 22-MAY-2000; 2000WO-US14042.
PR 05-JUN-2000; 2000US-0209832.
XX
(GETH ) GENENTECH INC.
XX
Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
XX
WPI; 2001-193260/18.
DR P-PSDB; AAB87560.
XX
Eighty four nucleic acids encoding PRO polypeptides, useful in
PI molecular biology, including use as hybridization probes, and in
PI chromosome and gene mapping.
XX
Claim 2; Fig 69; 278pp; English.
XX
The present sequence is the coding sequence for a human PRO polypeptide
CC (secreted and transmembrane). The PRO protein, and PRO agonists, PRO
CC antagonists or anti-PRO antibodies are useful for preparation of a
CC medicament useful in the treatment of a condition which is responsive to
CC the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO
CC protein may also be employed as molecular weight markers for protein
CC electrophoresis. The PRO coding sequence has applications in molecular
CC biology, including use as hybridisation probes, and in chromosome and
CC gene mapping.
XX
Sequence 3265 BP; 1159 A; 596 C; 632 G; 878 T; 0 other;
XX
alignment_scores:
Quality: 2879.50 Length: 914
Ratio: 3.779 Gaps: 8
Percent Similarity: 83.370 Percent Identity: 61.597
alignment_block:
US-09-049-696-41 x AAF92092
Align seg 1/1 to: AAF92092 from: 1 to: 3265
1. MetGlyProPheLysSerValPheIleLeuIleLeuHisLeuGlu1 17
25 ATGGGGTTATTTCAGAGGTTTGTTCCTTAGTTCTGCTGCTGCTGCA 74
19 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34
75 CCAGTCA...AATACCTCTTCATTAAAGCTGAATAATAATGGCTTTGAAG 121
34 lyIleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
122 ATATTGTCATTGTTATAGATCTTCAGTGTGCCAGAGATGAAAAATAATT 171
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
172 GAACAAATAGAGGATATGGTGACTACAGCTTCTACGTACCTGTTGAAGC 221
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluT 84
222 CACAGAAAAAAGATTTTTCAAAAATGTATCTATATTAATTCCCTGAGA 271
84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
272 ATTGGAGAGAAATCTCAGTCAAAAGGCCAAACATGAAACCATATAA 321
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
322 CATGCTGATGTTATAGTTGCACCACCTACACTCCAGGTAGAGATGAACC 371
117 cTyrThrGluGlnMetClyAsnCysGlyGluLysGlyGluArgIleHisL 134
372 ATACCAACCAAGCAGTTTACAGATGTGGAGAGAAAGGCCAATACATCACT 421
```

134 euThrProAspPheIleAlaGlyLysLeuAlaGluTyrGlyProGln 150
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
422 TCACCCCTGCAGCTTCTACTTGGAATAAAACAAAATGAATATGACCACCA 471

151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
472 GCACAACGTGTTGTCCATGAGTGGCGTCACCTCCGGTGGGAGTGTTTGA 521

167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGly...ArgIleG 183
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
522 TCAGTACAATCAAGATCAGCCCTTCTACCGTGCTAAGCTCAAAAAAATCG 571

183 InAlaValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLys 199
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
572 AAGCAACAAGTGTGCCGAGGTATCTCTGGTAGAATAAGATTATATAAG 621

200 CysGlnGlyGlySerCysTyrThrLysArgCysThrPheAsnLysValTh 216
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
622 TGTCAAGGAGCGAGCTGTCTTAGTAGAGCATGCAGAATTGATCTACAAC 671

216 rGlyLeuTyrGluLysGlyCysGluPheValLeuLnsrArgGlnThrG 233
| | |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
672 AAAACTGTATGGAAGAAGATTGTCAATTCTTCCCTGTGATAAAGTACAAACAG 721

233 LuLysAlaSerIleMetPheAlaGlnHisValaspSerIleValGluPhe 249
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
722 AAAGAAGCATCATATGTTTATGCAAGTAGTAATTGATCTGTGTGTAATTT 771

250 CysThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLy 266
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
772 TGTAACGAAAAAACCCATAATCAAGAAAGCTCCAAGCCCTACAAAACATAAA 821

266 SCysAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheL 283
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
822 GTGCAATTTTGAAGTACATGGGAGGTGATTAGCAATTCTGAGAGATTTTA 871

283 ySLysThrThrPromethThrThrGlnProProAsnProThrPheSerLeu 299
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
872 AAAACACCATCACCATGGTGACACCACTCCTCCACCTGCTCTCTCATTTG 921

300 LeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySe 316
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
922 CTGAAGATCAGTCAAAGAAATTTGTGTGTTAGTCTCTCATAGTCTGGAAG 971

316 rMetAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeup 333
||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
972 CATGGGGGGTAAAGGACCCCTTAATTCGATGAAATCAAGCAGCAAAACATT 1021

333 heLeuLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPhe 349
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1022 TCTCTCTGCAGACTCTTCAAAATGATCCTGGTGGGAGTGGTTCACTTT 1071

350 AspSerAlaAlaHisValGlnSerCluleuIleGlnIleasnSerGlySe 366
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1072 GATAGTACTGCCATTATGTTAAATAAGCTAATCCAAATAAAAGACGTGA 1121

366 rAspArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyT 383
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1122 TCAAGAAACACACTCATGGCAGGATTACCTACATATCCTCTGGGAGGAA 1171

383 hrSerIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLys 399
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1172 CPTCCATCTGCTGGAATTAATAATGCATTTCAGGTGATTGGAGACGCTA 1221

400 Tyr...ProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAs 415
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1222 CATTCCCAACACTCGATGGATCCGAGTACTGCTGCTGACTGNTGGSGAGGA 1271

415 psAnThrIleSerGlyCysPheAsnGluValLysGlnserGlyAlaIleI 432
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1272 TAACACTGCAAGTCTCTGATTGATGAAGTCAAGTCAAAAGTGGGCCATTG 1321

432 leHisThrValAlaLeuGlyProSerAlaLaGlnGluLeuGluGluLeu 448
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::
1322 TTCATTATTTATTCCTTTGGGAAGACGTCTGATGAAGCAGTAATAAGCATG 1371

449 SerLysMetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAs 465
||||| ::||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1372 AGCAAGAATAACAGAGGAAGTCATTTTATGTTTTCAGATGAAGCTCGAAG 1421

466 nAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlav 482
||||| ::||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1422 CAATGGCCCTCATTGATGCTTTTGGGGCTCTTACATCAGGAATACTGATC 1471

482 alSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsn 498
||||| ::||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1472 TCTCCAGAAAGTCCTTCAGCTCGAAAAGTAGGAGGATTAACACTGAATAGT 1521

499 SerGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLysAs 515
::: ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1522 AATGGCTCGATGAACGACACTGTCATATGATGATGATGACGGGGAAGGA 1571

515 pThrLeuPheLeuIleThrTrpThrGlnProGlnIleLeuLeu 532
::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1572 CAGCTTCTCTCATCATCGAACAGCTGCCCTCCAGATATTTCTCTCT 1621

532 rpAspProSerGlyGlnLysGlnGlyPheValValAspLysAsnThr 548
::: ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1622 GGGATCCCAAGTGAACAATAATGGAATAATTTCCAGTGGATGCACATCCC 1671

549 LysMetAlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLys 565
::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1672 AAAATGGCTATPCATGATTTTCCAGGAACCTCCAAGGTGGGCATCTGGGC 1721

565 stySerLeuGlnAlaSerSer.....GlnThrLeuThrLeuThrValT 580
::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1722 ATCAACTCTTCAGCCAAAGCGAACCCAGAACATTAACATTACAGTAA 1771

580 hrSerArgAlaSerAsnAlaThrLeuProProIleThrValThrSerLys 596
::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1772 CTTCGAGCAGCAAAATTCCTCTGTGCGCTCCAATCACAGTGAATGCTAAA 1821

597 ThrAsnLysAspThrSerLysPheProSerProLeuValValTyrAlaAs 613
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1822 ATGAATAAGGACGTAAACAGTTCCCCAGCCCAATGATGTTTACGCAGGA 1871

613 nIleArgGlnGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuI 630
::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1872 AATTCACAGGATATGTAACCTGTCTTGGAGCCCAATGTGACTGCTTCA 1921

630 leGluSerValAsnGlyLysThrValThrLeuGluLeuLeuAspAsnGly 646
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1922 TTGAATACAGAATGGACATACAGAAGTTTTTGGAACTTTTGGATAATGGT 1971

647 AlaGlyAlaAspAlaThrLysAspAspGlyValTyrSerArgTyrPheTh 663
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1972 CGACGCGTGATTCCTTCCAGAAGATGAGAGTCTACATCCAGGATATTTAC 2021

663 rThrTyrAspThrAsnGlyArgTyrSerValLysValArgAlaLeuGly 680
::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
2022 AGCATATACAGAAATGCGACATATAGCTTAAAGTTCCGGCTCATGGAG 2071

680 lyValAsnAlaAlaArgArgValIleProGlnGlnSerGlyAlaLeu 696
::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
2072 GAGCAAACACTGCGAGGCTAAAAATTACGGCTCCCACTCAATAGAGCCGCG 2121

697 TyrIleProGlyTyrIleGluAsnAspGluIleGlnTrpAsnProProAr 713
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2122 TACATACCAGCGTGGGTAGTGAACGGGGAATTTGAAGCAAAACCGCCCAAG 2171

713 qProGluIleAsnLysAspAspValGlnHisLysGlnValCysPheSerA 730
::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
2172 ACCTGAAATGTGAT...GAGGATACACAGCACACCTTGAGGAGATTTCAGCC 2218

730 rgThrSerSerGlyGlySerPheValAlaSerAspValProAsnAlaPro 746

122	ATATTGTCTCAATGTTTATAGATCCTACTGTGCCAGAGATGAAAAATAATT	171
51	GinGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl	67
172	GAACAAATAGAGGATATGGTGACTACAGCTTCTACGTACCTGTGTTGAAC	221
67	aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluT	84
222	CACAGAAAAAGATTTTTTCAAAAAATGATCTATATTAATTCCTGAGA	271
84	hTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys	100
272	ATTGGAAGAAATCCTCAGTACAAAGGCCAAACATGAAACCATATAA	321
101	AsnAlaAspValLeuValalacIuSerThrProGlyAsnAspGluPr	117
322	CATGCTGATGTTATAGTTGCACCACCTACACTCCAGGTAGAGATGAAC	371
117	oTyrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL	134
372	ATACACCAAGCAGTTCACAGAAATGTGGAGAGAAAGGCGAATACATTCAC	421
134	euThrProAspPheIleAlaGlyLysLeuAlaGluTyrGlyProGln	150
422	TCACCCCTGACCTCTACTTGGAAAAACAAATGAATATGGACCAACA	471
151	GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs	167
472	GSCAAACTGTTTGTCCATGAGTGGGCTCACCTCCGGTGGGAGGTGTTGA	521
167	pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGly...ArgIleG	183
522	TGAGTACAAATGAAGATCAGCCTTCTACCGTGCCTAAAGTCAAAAAATCG	571
183	lnAlaValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLys	199
572	AGCAACAAGGTGTTCCGCGAGTATCTCTGGTAGAAATAGAGTTTATAAG	621
200	CysGlnGlyGlySerCysTyrThrLysArgCysThrPheAsnLysValTh	216
622	TGTCAGGAGGCAGCTGCTTAGTAGTAGCATGCAGATTTGATTCACAA	671
216	rGlyLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrG	233
672	AAAACTGTATGCAAAAGATGTGCAATTTCTTCCTGATAAAGTACAAACAG	721
233	luLysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPhe	249
722	AAAAAGCATCCATAATGTTTATGCAAAAGTATGATCTGTTGTTGAATTT	771
250	CysThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLy	266
772	TGTACGAAAAAACCCATATACAGAGCTCCAGGCTCACAAACATATAA	821
266	sCysAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheL	283
822	GTGCAATTTTACAAGTACATGGAGGTGATTAGCAATTCGAGGATTTTA	871
283	ysLysThrThrProMetThrThrGlnProProAsnProThrPheSerLeu	299
872	AAACACCATACCATCGGTGACACACCTCCCTCCACCTGCTCTCTCATATG	921
300	LeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySe	316
922	CTGAGATCAGTCAAGAAATTTGTGTCTAGTTCTTGATAAGTCTGGAAG	971
316	rMetAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeup	333
972	CATGGGGGTAAAGGACCGCTAAATCGAATGAATCAACAGCAGCAAAACATT	1021
333	heLeuLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPhe	349
1022	TCCTGCTGCAGACTGTGAAATGATGATCCTGGGTGGGATGGTTTCACTTT	1071

350 AspSerAlaAlaHisValGlnSerGluLeuLeuLeuGlnIleAsnSerGlySer 365
 1072 GATAGTACTGCCACTATTGTTAAATAAGACTAATCCAAATAAAAAACCACTGA 1121
 366 rAspArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyT 383
 1122 TGAAAGAAACACACTCATGGCAGGATTACCTACATATCTCTCTGGGAGGAA 1171
 383 hrSerIleCysSerGlyLeuArgSerAlaPheThrValIleArgLeuLys 399
 1172 CTTCCATCTGCCTCGAATTAATAATGCAATTTCAAGTCAGTGTGGAGACCTA 1221
 400 Tyr...ProThrAspGlySerGluIleValLeuLeuLeuThrAspGlyGluAs 415
 1222 CATTCCCAACTCGATGGATCCGAGTACTGCTGTGACTGATGGGAGGA 1271
 415 pAsnThrIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleI 432
 1272 TAACACTCCAACTCTCTGATTGATGAAGTGAACAAGTGGGCCCATTTG 1321
 432 leHisThrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluGluLeu 448
 1322 TTCATTATTGCTTTGGGAAGAGCTGCTGATGAAGCAGTAATAGACATG 1371
 449 SerLysMetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAs 465
 1372 AGCAAGATAACAGGAGGAAGTCAATTTTATGCTTTTCAGATGAAGCTCAGAA 1421
 465 nAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaV 482
 1422 CAATGGCCTCATTTGATGCTTTTGGGGCTTTACATCAGGAATAACTACTGTC 1471
 482 aIserGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsn 498
 1472 TCTCCAGAAGTCCCTTCAGCTCGAAGTAAGGATTAACACTCAACTCAATAGT 1521
 499 SerGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLysAs 515
 1522 AATGCTGGATGACACACACTGTCATAATTGATAGTACAGTGGGAGGAA 1571
 515 pThrLeuPheLeuIleThrTrpThrGlnProProGlnIleLeuLeuT 532
 1572 CACGTCTCTTCATCACAATGAACAGCTCGCTCCCTCCCAAGTATTTCTCTCT 1621
 532 rPAspProSerGlyGlnGlyGlnGlyGlyPheValValAspLysAsnThr 548
 1622 GGGATCCCCAGTCGGAACAATAATAGAAAATTTACAGTGGATGCAACTTCC 1671
 549 LysMetAlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLys 565
 1672 AAAATGGCCTATCTCAGTATTTCCAGGAACCTGCAAAAGTGGGCACCTTGGGC 1721
 565 sTyrSerLeuGlnAlaSerSer.....GlnThrLeuThrLeuThrValT 580
 1722 ATACAATCTTCAAGCCAAAGCGAACCCAGAACCAATTAACATTACAGTAA 1771
 580 hrSerArgAlaSerAsnAlaThrLeuProProIleThrValThrSerLys 596
 1772 CTTCTCGAGACGCAAAATCTCTGTGCTCCCAATCACAAGTGAATGCTTAAA 1821
 597 ThrAsnLysAspThrSerLysPheProSerProLeuValValTyrAlaAs 613
 1822 ATGATATAGGACGTAAACAGTTTCCCACAGCCCAATGATTGTTTACGCAGA 1871
 613 nIleArgGlnGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuI 630
 1872 AATTCTACAAGCATATGTACCTGTTCTTGGAGCCCAATCTGACTGCTTCA 1921
 630 leGluSerValAsnGlyLysThrValThrLeuGluLeuLeuAspAsnGly 646
 1922 TTTCATTCACAAAGTGGACATACAGAAGTTTGTGGAATCTTTTGGATTAATGGT 1971

647 AlaGlyAlaAspAlaThrLysAspAspGlyValTyrSerArgTyrPheTh 663
|||||
1972 GCAGGCCCTGATTCTTCAAGAATGATGGAGTCTACTCCAGGTATTTTAC 2021
663 rThrTyrAspThrAsnGlyArgTyrSerValLysValArgAlaLeuGlyG 680
|||||
2022 AGCATATACAGAAATGGCAGATATAGCTTAAAGTTCCGGCTCATGGAG 2071
680 lyValAsnAlaAlaArgArgAlaValIleProGlnGlnSerGlyAlaLeu 696
|||||
2072 GAGCAAACTGCGCAGGCTAAAATACGGCTCCACTGAATAGAGCGCG 2121
697 TyrIleProGlyTrpIleGluAsnAspGluIleGlnTrpAsnProProAr 713
|||||
2122 TACATACCGGCTGGGTAGTGAACGGGAAATGAAGCAAAACCCGCCAAG 2171
713 gProGluIleAsnLysAspValGlnHisLysGlnValCysPheSerA 730
|||||
2172 ACCTGAAATTGAT...GAGGATACTCAGACCACCTTGGAGGATTTCAGCC 2218
730 rgThrSerSerGlyGlySerPheValAlaSerAspValProAsnAlaPro 746
|||||
2219 GAACAGCATCCGGAGGTGCTATTGTGTGATFACAAAGTCCCAAGCTTCCC 2268
747 IleProAspLeuPheProProGlyGlnIleThrAspLeuLysAlaGluIl 763
|||||
2269 TTGCTTGACCAATACCATCAACCAATCAACACAGACCTTGATGCCACAGT 2318
763 eHisGlyGlySerLeuIleAsnLeuThrTrpAlaProGlyAspAspt 780
|||||
2319 TCATGAGGATAAGATTATT...CTTACATGGACAGCACCAGGAGATAATT 2365
780 yrAspHisGlyThrAlaHisLysTyrIleIleArgIleSerThrSerIle 796
|||||
2366 TTGATGTGGAAAGTTCAACGTTATATCATATAAGATAAGTCCAAGTATT 2415
797 LeuAspLeuArgAspLysPheAsnGluSerLeuGlnValAsnThrThrAl 813
|||||
2416 CTTGATCTAAGACACAGTTTTGATGATGCTCTTCAAGTAAATACTACTGA 2465
813 aLeuIleProLysGluAlaAsnSerGluGluValPheLeuPheLysProG 830
|||||
2466 TCTGTCAACCAAGAGGCCAACTCCAAGGAAGCTTTGCATTTAAACCAG 2515
830 luAsnIleThrPheGluAsnGlyThrAspLeuPheIleAlaIleGlnAla 846
|||||
2516 AAATATCTCAGAGAAATGCAACCCACATATTATTGCCATTAAAGT 2565
847 ValAspLysValAspLeuLysSerGluIleSerAsnIleAlaArgValSe 863
|||||
2566 ATAGATAAAGCAATTGTGACATCAAAAGTATCCAACATTGCACAAGTAAC 2615
863 rLeuPheIleProProGlnThrProProGlu...ThrProSerProAspG 879
|||||
2616 TTTGTTTATCCCTCAAGCAAACTCTGATGACATTGATCTTACACCTACTC 2665
879 luThrSerAlaProCysProAsnIleHisIleAsnSerThrIleProGly 895
|||||
2666 CTACTCTCTACTCTCTCTGATAAAAGTCATAAATTCT.....GGA 2706
896 IleHisIleLeuLysIleMetTrpLysTrpIleGlyGluLeu 909
|||||
2707 GTTAATATTCTACGCTGGGTATTGTCTGTGATTGGGTCTGTT 2748

Tue Apr 2 09:40:05 2002

US-09-049-696-41 x AF127036

Align seg 1/1 to: AF127036 from: 1 to: 2826

1 MetGlyProPheLysSerValPheIleLeuLeuHisLeuLeuG1 17
5 ATGGGGCCATTAAAGAGTTCTGTGTTCATCTTGATCTTCACCTTCAGA 54
17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34
55 AGGGGCCCTGAGTAATTCATCTACCTGACCTGNACCAATGGCTATGAAG 104
34 lyIleValValAlaIleAspProAsnValProGluAspGluThrIleu 50
105 GCATTGCTGCTGCAATCGACCCCAATGTGCCAGAAAGATCAAACTCAT 154
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
155 CAACAATAAAGGACATGTGTGACCCAGGCATCTCTGTATCTGTTGAAGC 204
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGlu 84
205 TACAGGAAGCGATTTATTTCAAAATGTTGCCATTTTGATTCCTGAAA 254
84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
255 CATGGAGACAAAGGCTGATGTGTGAGACCAAACTTCAGACCTACAAA 304
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
305 AATGCTGATGTTCTGGTGTGAGTCTACTCTCCAGGTAATGATGAACC 354
117 oTyrThrGluGlnMetGlyAsnGlyGlyGluLysGluGluGluIleHis 134
355 CTACACTGACGAGATGGCAACTGTGGAGAGAGGGGTGAAGGATCCACC 404
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
405 TCACCTCTGATTTTCATTCAGGAAAAAGTTAGTGAATATGGACCACAA 454
151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
455 GGTAGGGCATTTGCCATGAGTGGGCTCATCTACGATGGGGAGTATTGA 504
167 pClyTrpAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGln 184
505 CGAGTACAAATGATGAGAAATCTACTTATCAATGGAAGATACAAG 554
184 lavalArgCysSerAlaGlyIleThrGlyThrAsnValValLysLysCys 200
555 CAGTAAGATGTTACGACGGTATTCTGGTACAAATGTAGTAAAGATGT 604
201 GlnGlyGlySerCysTyrThrLysArgCysThrPheAsnLysValThrG 217
605 CAGGGAGCAGCTGTACACCAAAAGATGCACATTCATAAAGTTACAGG 654
217 yLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrGlu 234
655 ACTCTATGAAAAGGATGTGAGTTGTTCTCCAAATCCCGCCAGCGGAGA 704
234 ysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPheCys 250
705 AGGCTTCTAATATGTTGGCAACATGTTGATTCATATGATTGAATTCGT 754
251 ThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLysCys 267
755 ACAGACAAACCAACAAAGAGCTCCAAACAGCAAAATCAAAAATG 804
267 sasnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheLysL 284
805 CAATCTCCGAGCACATGGGAGTGTGCTGATCTGAGGACTTTAAGA 854
284 ysthrThrProMetThrThrGlnProProAsnProThrPheSerLeuLeu 300

855 AACCACCTCTATGACACACAGCCACCAAAATCCACCTTCTCATTTGCTG 904
301 GlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySerMe 317
905 CAGATTGGACAAAGAAATGTGTGTAGCTTGACAAATCTGGGAAGCAT 954
317 tAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheL 334
955 GGGGACTGGTAACCGCTCAATCGACTGAATCAAGCAGGCCAGCTTTTCC 1004
334 euLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPheAsp 350
1605 TGCTGCAGACAGTTGAGCTGGGTCTGGGTGGGATGGTACATTGGAC 1054
351 SerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySerAs 367
1055 AGTGCTGCCCATGTACAAAGTGAATCATACAGATAACAGTGCAGTGA 1104
367 pArgAspThrLeuAlaLysArgLeuProAlaAlaSerGlyGlyThrS 384
1105 CAGGCACACACTGCCCAAAAGATTTACCTGCAGCAGCTTCAGGAGGACGT 1154
384 erIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLysTyr 400
1155 CCATCTGCAGGGGCTTCGATCGGCATTTACTGTGATTAGGAAGAAATAT 1204
401 ProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAsnTh 417
1205 CCAACTGATGATCTGAAATGTTGCTGTACGGATGGGAAGACACAC 1254
417 rIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleHist 434
1255 TATAAGTGGGTGCTTTACGAGGTCACAAAGTGGTCCATCATCCACA 1304
434 hrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluGluLeuSerLys 450
1305 CAGTGCCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGAGCTGTCCAAA 1354
451 MetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnAsnG 467
1355 ATGACAGGAGGTTTACAGACATATGCTTCAGATCANGTTTCAGAACATGG 1404
467 yLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValSerG 484
1405 CCTCATGATGCTTTTGGGGCCCTTTCATCAGGAAATGGAGCTGTCTCTC 1454
484 lArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSerGln 500
1455 AGCGCTCCATCCAGCTTGAGAGTAAAGGATTAACCTCCAGAACAGCCAG 1504
501 TrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspThrIle 517
1505 TGGATGAATGGCACAGTGTGTCGGACAGCAGCTGGGAAAGGACACTTT 1554
517 uPheLeuIleThrTrpThrThrGlnProProGlnIleLeuLeuTrpAspP 534
1555 GTTTCCTTATCACCTGGACACAGCGCTCCCAAAATCTCTCTCTGGATC 1604
534 roSerGlyGlnLysGlnGlyPheValValAspLysAsnThrLysMet 550
1605 CCAGTGGACAGAACGAGGTGGCTTTGTAGTGACAAAAACCAAAATG 1654
551 AlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLysTyrSe 567
1655 GCCTACCTCCAAATCCAGGCATTTGTTAGGTGGCATTGGGAATACAG 1704
567 rLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSerArgAlaS 584
1705 TCTGCAAGCAAGCTCACAAACCTTGACCTGACTGTACGTCGTCGCGT 1754
584 erAsnAlaThrLeuProProIleThrValThrSerLysThrAsnLysAsp 600


```
1755 CCAATGCTACCCCTCCCTCAATTACAGTGACTTCCAAAACGAACAAGGAC 1804
601 ThrSerLysPheProSerProLeuValValTyrAlaAsnIleArgGlnGI 617
1805 ACCAGCAAAATTCGCCACCCCTCTGGTAGTTATGCAAAATATTCGCCAAGG 1854
617 yAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerValA 634
1855 AGCTTCCCAATTTCTAGGCCAGTGTCACAGCCCTGATTGAATCAGTGA 1904
634 snGlyLysThrValThrLeuCluLeuLeuAspAsnGlyAlaGlyAlaAsp 650
1905 ATGGAACAAACAGTTACCTTGGAACTACTGGATATGAGCAGGTGCTGAT 1954
651 AlaThrLysAspGlyValTyrSerArgTyrPheThrThrTyrAspTh 667
1955 GCTACTAAGGATGACGGTGCTCTACTCAAGGTATTTTCAAACTTATGACAC 2004
667 rAsnGlyArgTyrSerValLysValArgAlaLeuGlyGlyValAsnAlaA 684
2005 GAATGGTAGATACAGTGTAAAGTGGCGGCTCTGGAGGAGTTAAACGACG 2054
684 laArgArgValIleProGlnGlnSerGlyAlaLeuTyrIleProGly 700
2055 CCAGCGAGAGTGATACCCAGCAGAGTGGAGCAGCTGTATACATCCTGGC 2104
701 TrpIleGluAsnAspGluIleGlnTrpAsnProProArgProGluIleAs 717
2105 TGGATTGAGAAATGATGAATACAAATGGAATCCACCAAGAGCTGAAATTA 2154
717 nLysAspValGlnHisLysGlnValCysPheSerArgThrSerSerG 734
2155 TAAGGATGATGTCAACACAAAGCAAGTGTGTTTCAGCAGAACATCCTCGG 2204
734 lyGlySerPheValAlaSerAspValProAsnAlaProIleProAspLeu 750
2205 GAGGCTCATTTGTGGCTCTGATGTCCTCCAAATGCTCCATACCTGATCTC 2254
751 PheProGlyGlnIleThrAspLeuLysAlaGluIleHisGlyGlySe 767
2255 TTCCACCTCGCCAAATCACCGACCTGAAGCGGAAATTCACGGGGCAG 2304
767 rLeuIleAsnLeuThrTrpThrAlaProGlyAspTyrAspHisGlyT 784
2305 TCTCATTAATCTGACTTGGACAGCTCTCTGGGATGATATGACCATGGAA 2354
784 hrAlaHisLysTyrIleIleArgIleSerThrSerIleLeuAspLeuArg 800
2355 CAGCTCACAAAGTATATCATTCGAATAAGTACAAGTATTTCTGATCTCAGA 2404
801 AspiLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuIleProLy 817
2405 GACAAGTTCAATGAATCTCTTCAAGTGAATACTACTGCTCTCATCCAAA 2454
817 sGluAlaAspSerGluGluValPheLeuPheLysProGluAsnIleThrP 834
2455 GGAAGCCAACTCTGAGGAGTCTTTTGTGTTTAAACCAAGAAACATTAATT 2504
834 heGluAsnGlyThrAspLeuPheIleAlaIleGlnAlaValAspLysVal 850
2505 TTGAAAATGGCACAGATCTTTTTCATTGCTATTCAGGCTGTTGATAAGGTC 2554
851 AspleuLysSerGluIleSerAsnIleAlaArgValSerLeuPheIlePr 867
2555 GATCTGAAATCAGAAATATCCAACTTGCACGAGTATCTTTGTTTATATCC 2604
867 oProGlnThrProProGluThrProSerProAspGluThrSerAlaProC 884
2605 TCCACAGACTCCGCCAGACACCTAGTCTCTGATGAACGCTGCTCTCTT 2654
884 ysProAsnIleHisIleAsnSerThrIleProGlyIleHisIleLeuLys 900
2655 GTCTTAATATTCATATCAACAGCACCATTCTTGGCATTCACATTTTAA 2704
```

```
901 IleMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914
2705 ATTATGTGAAGTGGATAGGAAACTGCAGCTGCAATAGCC 2746
```

seq_name: gb_pat:AX193489

seq_documentation_block: 3311 bp DNA PAT 15-AUG-2001
LOCUS AX193489 Sequence 1056 from Patent WO0149716.

ACCESSION AX193489

VERSION AX193489.1 GI:15211440

KEYWORDS human:

SOURCE

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 3311)

AUTHORS Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J.,

Stolk,J.A., King,G.E., Wang,T. and Jiang,Y.

Compounds for immunotherapy and diagnosis of colon cancer and

methods for their use

PATENT: WO 0149716-A 1056 12-JUL-2001;

CORIXA CORPORATION (US)

FEATURES Location/Qualifiers

source 1..3311

organism "Homo sapiens"

db_xref="taxon:9606"

BASE COUNT 1028 a 692 c 742 g 849 t

ORIGIN

alignment_scores:

Quality: 4756.00 Length: 914

Ratio: 5.204 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.891

alignment_block:

US-09-049-696-41.x AX193489 ..

Align seg 1/1 to: AX193489 from: 1 to: 3311

MetGlyProPheLysSerSerValPheIleLeuIleLeuHisLeuLeuGI 17

ATGGGGCCATTTAAGAGTCTGTGTTCATCTGATTTCTACCTTCTTAGA 401

uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34

AGGGGCCCTGAGTAATTCACCTCATTCAGCTGAACAACANTGGCTATGA 451

lyIleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50

GCATTTGCTTGCATCGACCCCAATGTGCCAGAGATGAAACACTCATT 501

GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67

CAACAAATAAAGGACATGGTGACCCAGCATCTCTGTATCTGTTTGAAGC 551

aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluT 84

TACAGGAAGCGATTTATTTCAAAAATGTTGCCATTTTGATTCCTGAAA 601

hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100

CATGGAAGACAAAGGCTGACTATGTGACACCAAACTTGAGACCTTACAAA 651

AsnAlaAspValLeuValAlaGluSerThrProProGlyAsnAspGluPr 117

AATCCTGATGTTCTGGTGTGCTGAGTCTACTCCTCAGGATGATGAACC 701

oTyrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134

CTACACTGAGCAGATGGCAACTGTGGAGAGAGGGTGAAGGATCCACC 751

134 euThrProaspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
|||||
752 TCACCTCCTGATTTTCATGCGAGGAAAAAGTTAGCTGAATATGACCACAA 801

151 GlyArgAlaPheValHisGluTyrAlaHisLeuArgTrpGlyValPheAs 167
|||||
802 GGTAAGGCATTTCCTCATGATGATGGGCTCATCTACGATGGGATATTGA 851

167 pGluTyrAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGlnA 184
|||||
852 CGAGTACAATAATGATGAGAAATTCATTTATCCAATGGAAGAAATACAAG 901

184 laValArgCysSerAlaGlyIleThrGlyThrAsnValLysLysCys 200
|||||
902 CAGTAAGATGTTTCAGCAGGATTTACTGGTACAAATGCTAGTAAGAAGTGT 951

201 GlnGlySerCysTyrThrLysArgCysThrPheAsnLysValThrG 217
|||||
952 CAGGAGGCAGCTGTTACACCAAAAGATGCACATTCATTAAGTTACAGG 1001

217 yLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrGlu 234
|||||
1002 ACTCTATGAAAAGGATGTGAGTTTCTCCAAATCCCGCCAGACGAGA 1051

234 ysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPheCys 250
|||||
1052 AGGCTTCTATATGTTTGGCAACATGTTGATCTCATAGTTGAATTCGT 1101

251 ThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLysCy 267
|||||
1102 ACAGAACAAACACACAAACAAAGCTCCAAACAAAGCAAAATCAAAAATG 1151

267 sAsnLeuArgSerThrTrpGluValIleArgspSerGluAspPheLys 284
|||||
1152 CAATCTCCGAGACATGGGAAGTATCGTATTCTGAGGACTTTAAGA 1201

284 ysThrThrProMetThrThrGlnProProAsnProThrPheSerLeuLeu 300
|||||
1202 AAACCACTCCTATGACACACACACGCCAAATCCCACTTCTCATTTGCTG 1251

301 GlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySerMe 317
|||||
1252 CAGATTGGACAAAGAAATGTGTGTTTACTCTTGACAAATCTGGAAGCAT 1301

317 tAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheL 334
|||||
1302 GSCGACTGGTAACCCCTCAATCGACTGATCAAGCAGGCCAGCTTTCC 1351

334 euLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPheAsp 350
|||||
1352 TGCTGCAGACAGTTGAGCTGGGGTCTGGTGGGATGGTGCATTTGAC 1401

351 SerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySerAs 367
|||||
1402 AGTGCTGCCATGTRCAAGTGAACTCATACAGATAAAACAGTGGCAGTGA 1451

367 pArgAspThrLeuAlaTyrArgLeuProAlaAlaAlaSerGlyGlyThrS 384
|||||
1452 CAGGACACACTCGCCAAAAGATTACCTGCAGCAGCTTCAGGAGGAGCT 1501

384 erIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLysTyr 400
|||||
1502 CCATCTGCAGCGGGCTTCGATCGGCATTTACTGTGATTAGGAAGAAATAT 1551

401 ProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAsnTh 417
|||||
1552 CCACTGATGATCTGAATTTGCTGCTGACGGATGGGAAGACAAAC 1601

417 rIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleHisT 434
|||||
1602 TATAAGTGGGTGCTTTACGAGGTCAAACAAAGTGGTCCCATCATCCACA 1651

434 hrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluLeuSerLys 450
|||||
1652 CACTCGCTTTGGGGCCCTCTGACAGCTCAAGAACTAGAGGAGCTGCCAAA 1701

451 MetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnGln 467
|||||
1702 ATGACAGGAGGTTTACAGACATATGTTTCAGATCAAGTTTCAGAACAAATGG 1751

467 yLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValSerG 484
|||||
1752 CCTCATGTGATGCTTTTGGGCCCTTTTCATCAGGAATGGAGCTGTCTCTC 1801

484 laArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSerGln 500
|||||
1802 AGCCCTCCATCCAGCTTGAGGTGAGGTAAACCCCTCCAGAACACGCCAG 1851

501 TrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspThrLe 517
|||||
1852 TGGATGAATGGGCACAGTGTGCTGGGACAGCACCGTGGGAAAGGACACTTT 1901

517 upheLeuIleThrTrpThrThrGlnProProGlnIleLeuLeuTrpAspP 534
|||||
1902 GTTTCCTTATCACTGGACACAGCGCTCCCAAAATCCTTCTCTGGGATC 1951

534 roSerGlyGlnLysGlnGlyGlyPheValValAspLysAsnThrLysMet 550
|||||
1952 CCAGTGGACAGAAAGGAGGTGGCTTTAGTGGACAAAAACCAAAATG 2001

551 AlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLysTyrSe 567
|||||
2002 GCCTACCTCCAATCCAGGCAATGCTAAGGTGGCACCTTGGAAATACAG 2051

567 rLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSerArgAlas 584
|||||
2052 TCTGCAAGCAAGCTCACAAACCTTGACCTGACTGTACGTCCCGTCCGT 2101

584 erAsnAlaThrLeuProProIleThrValThrSerLysThrAsnLysAsp 600
|||||
2102 CCAATGCTACCCCTGCCCTCCAATTACAGTACTTCCAAAAACGAACAGGAC 2151

601 ThrSerLysPheProSerProLeuValValTyrAlaAsnIleArgGlnG 617
|||||
2152 ACCAGCAAAATCCCGAGCCCTCTGGTAGTTTATGCAAAATATTCGCCAAG 2201

617 yAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerVala 634
|||||
2202 AGCCTCCCAATTCAGGGCCAGTGTACAGCCCTGATTGAATCAGTGA 2251

634 snGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAlaAsp 650
|||||
2252 ATGGAAAAACAGTTACCTTGGAACTACTGGATTAATGGACAGGTGCTGAT 2301

651 AlaThrLysAspAspGlyValTyrSerArgTyrPheThrThrTyrAspTh 667
|||||
2302 GCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCACAACTTATGACAC 2351

667 rAsnGlyArgTyrSerValLysValArgAlaLeuGlyGlyValAlaAsnAla 684
|||||
2352 GAATGGTAGATACAGTAAAGATCGGGCTCTGGGAGGAGTTAACGCAG 2401

684 laArgArgArgValIleProGlnGlnSerGlyValAlaLeuTyrIleProGly 700
|||||
2402 CCAGACGAGAGTGATACCCAGCAGAGTGGGAGCACTGTACATACCTGGC 2451

701 TrpIleGluAsnAspGluIleGlnTrpAsnProProArgProGluIleAs 717
|||||
2452 TGGATTGAGAATGATGAAATACAATGGAATCCACCAAGACCTGAAATTA 2501

717 nLysAspAspValGlnHisLysGlnValLysPheSerArgThrSerSerG 734
|||||
2502 TAAGATGATGTTCAACCAAGCAAGTGTGTTTCAGCAGAAACATCCTCGG 2551

734 lyGlySerPheValAlaSerAspValProAsnAlaProIleProAspLeu 750

||||| 2552 GAGGCTCATTTGGCTTCTGATGCCAAATGCTCCCATCTGATCTC 2601
751 PheProGlyGlnIleThrAspLeuLysAlaGluIleHisGlyGlyse 767
||||| 2602 TTCCACCTGCCAAATCACCACCTGAAGCGGAAATTCACGGGGGCGAG 2651
767 rLeuIleAsnLeuThrTrpThrAlaProGlyAspAspTyrAspHisGlyT 784
2652 TCTCATTAATCTGACTTGGACAGCTCTCGGGGATGATTATGACCATGAA 2701
784 hrAlaHisLysTyrIleIleArgIleSerThrSerIleLeuAspLeuArg 800
2702 CAGCTCACAAGTATATCATTCGAATAAGTACAAGTATTTTGTATCTCAGA 2751
801 AspLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuIleProLy 817
2752 GACAAGTTCATGAATCTCTTCAAGTGAATACTACTGCTCTCATCCCAA 2801
817 sGluAlaAsnSerGluGluValPheLeuPheIysProGluAsnIleThrP 834
2802 GGAAGCCAACTCTGAGGAAGTCTTTTGTGTTTAAACCCAGAAACATTA 2851
834 heGluAsnGlyThrAspLeuPheIleAlaIleGlnAlaValAspLysVal 850
2852 TTGAAAATGGCACAGATCTTTTCATGCTATTCAGGCTGTGATAAGGTC 2901
851 AspLeuLysSerGluIleSerAsnIleAlaArgValSerLeuPheIlePr 867
2902 GATCTGAAATCAGAAATATCCAAACATTCGACGAGTATCTTTGTTTATTC 2951
867 oProGlnThrProProGluThrProSerProAspGluThrSerAlaProC 884
2952 TCCACAGACTCCGCCAGACAGACCTAGTCCCTGATGAAACGCTGCTGCT 3001
884 ysProAsnIleHisIleAsnSerThrIleProGlyIleHisIleLeuLys 900
3002 GTCTAATATTCATATCAACAGCACCATTCTGCGATTCACATTTTAA 3051
901 IleMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914
3052 ATTATGTGGAAGTGATAGGAGAACTGACGCTGCAATAGCC 3093

seq_name: gb_pr:AF039400
seq_documentation_block:
LOCUS AF039400 3311 bp mRNA PRI 14-DEC-1998
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA, complete cds.
ACCESSION AF039400
VERSION AF039400.1 GI:4009457
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3311)
Gruber,A.D., Elble,R.C., Ji,H.L., Schreur,K.D., Fuller,C.M. and Pauli,B.U.
TITLE Genomic cloning, molecular characterization, and functional analysis of human CLCA1, the first human member of the family of Ca2+-activated Cl- channel proteins
Genomics 54 (2), 200-214 (1998)
MEDLINE 99047526
2 (bases 1 to 3311)
Gruber,A.D., Elble,R. and Pauli,B.U.
AUTHORS Direct Submission
TITLE Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA
JOURNAL
FEATURES
1. .3311
/organism="Homo sapiens"
/db_xref="taxon:9606"
source

/chromosome="1"
/map="lp22-p31"
/tissue_type="small intestine"
1. .3311
/gene="hCLCA1"
352. .3096
/gene="hCLCA1"
/note="transmembrane glycoprotein"
/codon_start=1
/product="calcium-dependent chloride channel-1"
/protein_id="AAC95428.1"
/db_xref="GI:4009458"
/translation="MGPFKSSVFILILHLEGALSNSLIQNNNGYEGIVVADPNVP
EDTLIQIKDMVTQASLYLFATGKRFYKFNVAIIPETWTKADYVRPKLETYKNA
DLVAESTPGNDPEYTEQNGCKGERIHLPTDFEIAKKLAEGYFGQKAFVHEWAH
LRMGVEFVNDEKFKYLSNGRIQACSGAGITGVNVKKGCGGCTTKRCKNTKWTGL
YEKCEFLQSRQTERKASIMFAQHVDSIVFECTEONHNKEAPNOKNOKNLRSTWEVI
RQSEDFKKTTPMTTOPNPFFSLQIQORLVCLVLDKSGSMATGNRLNRLNQAGQLFL
LQTVELGSWGVWTFDSAAHVQSELQINSNDRDLAKRLPAAAGGTSICGLRSA
FTVIRKKYPTDGSEI VLLTDGENTISGCFNEVKQSGAI IHTVALGSAAGLEELSLK
MTGGLQYASQVONNGLIDAFGALSSGNGAVSORSIQLESKGLITQNSOMNGTIV
DSTVGKDTLFLITWTPQPIILLWDPGQGGFVVDKNTKMYLQIPGIAKVGTKY
SLOASSOTLTLTVTSRASNATLPITVTSKNTKDTSKFPLVYVIANIROGASPLIRA
SVTALIESVNGKTVTLELLDNGAGADATKDGYSRYFTTYDNGRYSVAVRALGGVN
AARRVPIQSGALYIPGWTENDEIOWNPPEINKDDVOHKQVCFSRSTSGGSFVAS
DVPNAPIDLPFPQGQITDKAEIHGSLINLTWAPGDDYDGHGTAHKYIIIRISTILD
LRDKFNESLQVNTALIPKAEANSEEVFLFENITFENGTDLEIAIOAVKDLKSEI
SNIAVSLFIPPTPPTSPDETSAPCPNIHINSTIPGIIHILKIMKMWIGELQLSIA
"
BASE COUNT 1028 a 692 c 742 g 849 t
ORIGIN
alignment_scores:
Quality: 4756.00 Length: 914
Ratio: 5.204 Caps: 0
Percent Similarity: 100.000 Percent Identity: 99.891
alignment_block:
US-09-049-696-41 x AF039400
Align seg 1/1 to: AF039400 from: 1 to: 3311

1 MetGlyProPheLysSerSerValPheIleLeuIleLeuHisLeuLeuGl 17
352 ATGGGGCCATTAAGAGTCTGTGTCATCTTGATTTCTTCACCTCTTAGA 401
17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34
402 AGGGCCCTGAGTAATTCATCTCAGCTGAAACCAATGCTATGAAG 451
34 lyIleValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
452 GCATTTGCTTGCATATGACCCCAATGTCGCCAAGAGATGAAACACTCAT 501
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
502 CAACAAATAAAGGACATGGTACCAGGCATCTCTGTATCTGTTTGAAGC 551
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluT 84
552 TACAGGAAGCGATTATTTCAAAAATGTTGCCATTTTGATTCCTGAAA 601
84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
602 CATGGAACACAAAGGCTGACTATGTGACACCAAACTTGAGACCTACAAA 651
101 AsnAlaAspValLeuValAlaGluSerThrProProGlyAsnAspGluPr 117
652 AATGCTGATGTTCTGTTGCTGAGTCTACTCTCCAGGTAATGATGAACC 701
117 oTyrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134

702 CTACACTGAGCAGATGGGCAACTGTGGAGAGAAAGGTGAAAGGATCCACC 751
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
752 TCACCTCCTGATTGTTTCATTTGAGGAAAAAGTTAGCTGAATATGGACCACAA 801
151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
802 GGTAAAGGCAATTTGCCATGAGTGGGCTCATCTACGATGGGAGTATTGGA 851
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGlnA 184
852 CGAGTACAATAATGATGAGAAATTTCTACTTATCCAATGGGAATAACAAG 901
184 laValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLysCys 200
902 CAGTAAAGATGTTTCAGCAGGATTTACTGCTACAAATGTTAGTAAAGAGTGT 951
201 GlnGlyGlySerCysTyrThrLysArgCysThrPheAsnLysValThrG 217
952 CAGGAGGCAGCTGTTACACCAAAAGATGCACATTTCAATAAAGTTACAGG 1001
217 YLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrGluL 234
1002 ACTCTATGAAAAAGGATGTGAGTTTGTCTCCAATCCCGCCAGCAGGAGA 1051
234 ysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPheCys 250
1052 AGCTTCTATTAATGTTTGCACAAATGTTGATTCTATAGTTGAATTTCTGT 1101
251 ThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLysCy 267
1102 ACAGAACAAAAACCAACAAAGAGCTCCAAACAAACAAATCAAAAATG 1151
267 sasnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheLysL 284
1152 CAATCTCCGAAGCACATGGAAGTGAATCGTGATTCAGGACTTTTAAAGA 1201
284 ysThrThrProMetThrThrGlnProProAsnProThrPheSerLeuLeu 300
1202 AAACCACTCCTATGACACACACAGCCACCAAAATCCACCTTCATTTGCTG 1251
301 GlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySerMe 317
1252 CAGATTGGACAAGAAATGTTGTGTTAGTCTTGCATTAATCTGGAAGCAT 1301
317 tAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheL 334
1302 GCGACTGTGTAACCGCTCAATCGACTGAATCAACGAGCCAGCTTTTCC 1351
334 euLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPheAsp 350
1352 TGCTGCAGACAGTTGAGCTGGGCTCTGGGTGGGATGGTGACATTTGAC 1401
351 SerAlaAlaHisValGlnSerLeuLeuIleGlnIleAsnSerGlySerAs 367
1402 AGTGTGCCATGTACAAGTGAAGTCAATACAGATAAAACAGTGGCAGTCA 1451
367 pArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyThrS 384
1452 CAGGACACACTCGCAAAAGATTACCTGCAGCAGCTTCAGAGGGAGCT 1501
384 erIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLysTyr 400
1502 CCATCTGCAGGGGGCTTCGATCGCATTTACTGTGATTAGGAAGAAATAT 1551
401 ProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAsnTh 417
1552 CCAACTGATGATCTGAATTTGCTGCTGCTGACGATGGGGAAGACACAC 1601
417 rIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleHisT 434
1602 TATAAGTGGGTGCTTTAACGAGGTCAACAAAGTGGTGGCCATCATCCACA 1651

434 hrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluGluLeuSerLys 450
1652 CAGTCGCTTTTGGGGCCCTCTCCAGCTCAAGAACTAGAGAGCTGTCCAAA 1701
451 MetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnGln 467
1702 ATGACAGGAGTTTACAGACATATGCTTCAGATCAAGTTTCAGAACAAATGG 1751
467 YLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValSerG 484
1752 CCTCATTTGATGCTTTTGGGGCCCTTTCATCAGGAATGGAGCTGTCTCTC 1801
484 lnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSerGln 500
1802 AGCGCTCCATCCAGCTTGAGAGTAAGGATTAACCCCTCCAGAACACGCCAG 1851
501 TrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspThrLe 517
1852 TGGATGAATGGCACAGTGTGCTGGACAGCCGTGGAAAGGACACTTT 1901
517 uPheLeuIleThrTrpThrThrGlnProProGlnIleLeuLeuTrpAspP 534
1902 GTTTCCTTATCAGCTGACAAACGAGCCTCCCAAAATCCTTCTCTGGGATC 1951
534 roSerGlyGlnLysGlnGlyGlyPheValValAspLysAsnThrLysMet 550
1952 CCAGTGGACAGAACAGGTGGCTTTGTAGTGACAAAACACCAAAATG 2001
551 AlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLysTyrSe 567
2002 GCCTACTCCAAATCCAGGCATTTGTAAGTTGGCAGCTTGAATATACAG 2051
567 rLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSerArgAlaAs 584
2052 TCTGCAAGCAAGCTCACAAACCTTGACCTGACTGTACGCTCCCGTGGGT 2101
584 exAsnAlaThrLeuProProIleThrValThrSerLysThrAsnLysAsp 600
2102 CCAATGCTACCTGCTCCAAATTACAGTGACTTCCAAAACGAAACAGGAC 2151
601 ThrSerLysPheProSerProLeuValValTyrAlaAsnIleArgGlnG 617
2152 ACCAGCAATTTCCCGAGCCCTCTGGTAGTTTATGCAATATATTCGCCAAG 2201
617 yAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerValA 634
2202 AGCCTCCCAATTTCTCAGGGCCAGTGTACAGCCCTGATTGAATCAGTGA 2251
634 snGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAlaAsp 650
2252 ATGGAAAAACAGTTTACCCTTGGAACTACTGGATAATGGAGCAGGTCTGAT 2301
651 AlaThrLysAspAspGlyValTyrSerArgTyrPheThrThrTyrAspTh 667
2302 GCTACTAAGGATGACGGTGTCTACTCAAGGTATTTTCAACACTTATGACAC 2351
667 rAsnGlyArgTyrSerValLysValArgAlaLeuGlyGlyValAsnAlaA 684
2352 GAATGGTAGATACAGTGTAAAGTGGCGGCTCTGGAGAGGATTAACCCAG 2401
684 laArgArgArgValIleProGlnGlnSerGlyAlaLeuTyrIleProGly 700
2402 CCACCGGAGATGATACCCCGAGCAGAGTGGAGCAGTGTACATACCTGGC 2451
701 TrpIleGluAsnAspGluIleGlnTrpAsnProProArgProGluIleAs 717
2452 TGGATTGAGAATGATGAATACAATGAATCCACCAAGACCTGAAATTA 2501
717 nLysAspAspValGlnHisLysGlnValCysPheSerArgThrSerSerG 734
2502 TAAGGATGATGTTCAACACAAAGCAAGTGTGTTTTCAGCAGAACATCTCTCG 2551

```

734  lYgLYSerPheValAlaSerAspValProAsnAlaProIleProAspLeu 750
|||||
2552 GAGGCTCATTTGGTGGCTTCATGATGTCCTCCCAATGCTCCCATACCTGATCTC 2601
|||||
751  PheProGlyGlnIleThrAspLeuLysAlaGluIleHisGlyGlySe 767
|||||
2602 TTCCACCTGGCCANATCACCGACCTGAAGCGGAAATTCACGGGGCAG 2651
|||||
767  rLeuIleAsnLeuThrTrpThrAlaProGlyAspAspTyrAspHisGlyT 784
|||||
2652 TCTCATTAATCTGACTTGGACAGCTCTCGGGATGATATGACCATGAA 2701
|||||
784  hrAlaHisLysTyrIleIleArgIleSerThrSerIleLeuAspLeuArg 800
|||||
2702 CAGCTCAAGTATATCATTTGCAAGTAAAGTACAAATATCTTGATCTCAGA 2751
|||||
801  AspLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuIleProLy 817
|||||
2752 GACAAGTCAATGAATCTCTTCAAGTGAATACTACTCTCTCATCCCAA 2801
|||||
817  sGluAlaAsnSerGluGluValPheLeuPheLysProGluAsnIleThrP 834
|||||
2802 CGAAGCCAACCTCTCAGGAAGCTCTTTTGTGTTAAACCAAGAAACATTA 2851
|||||
834  heGluAsnGlyThrAspLeuPheIleAlaIleGlnAlaValAspLysVal 850
|||||
2852 TTGAAAATGCGACAGATCTTTTCAATGCTATTCAGGCTGTGATAAGGC 2901
|||||
851  AspLeuLysSerGluIleSerAsnIleAlaArgValSerLeuPheIlePr 867
|||||
2902 GATCTGAATCAGAAATATCCACATTCGACGAGTATCTTTGTTATTC 2951
|||||
867  oProGlnThrProGluThrProSerProAspGluThrSerAlaProC 884
|||||
2952 TCCACAGACTCCGCCAGACACCTAGTCTCTGATGAAACGCTGCTCCTT 3001
|||||
884  ysProAsnIleHisIleAsnSerThrIleProGlyIleHisIleLeuLys 900
|||||
3002 GTCTTATATTCATATCAACAGACACCATCTCTGGCATTCACATTTTAAA 3051
|||||
901  ileMetTrpLysTrpIleGlyGluLeuGlnLeuSerIleAla 914
|||||
3052 ATTATGTGAAGTGATAGAGAACTGCAGCTGTCAATAGCC 3093
|||||

```

seq_name: gb_om:AF095584

```

seq_documentation_block:
LOCUS AF095584 3079 bp mRNA MAM 13-OCT-2000
DEFINITION Sus scrofa epithelial chloride channel protein (AECC) mRNA,
complete cds.
ACCESSION AF095584
VERSION AF095584.1 GI:6002645
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactylia; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 3079)
AUTHORS Gaspar,K.J., Racette,K.J., Gordon,J.R., Loewen,M.E. and
Forsyth,G.W.
TITLE Cloning a chloride conductance mediator from the apical membrane of
porcine ileal enterocytes
JOURNAL Physiol. Genomics (Online) 3 (2), 101-111 (2000)
MEDLINE 20473747
PUBMED 11015605
REFERENCE 2 (bases 1 to 3079)
AUTHORS Gaspar,K.J., Gabriel,S.E., Racette,K.J. and Forsyth,G.W.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-1998) Veterinary Physiological Sciences,
University of Saskatchewan, 52 Campus Drive, Saskatoon, SK S7N 5B4,
Canada
FEATURES
source Location/Qualifiers
1..3079

```

```

/organism="Sus scrofa"
/db_xref="taxon:9823"
/tissue_type="ileal mucosa"
1..3079
/gene="AECC"
133..2886
/gene="AECC"
/function="involved in exocrine chloride conductance"
/codon_start=1
/product="epithelial chloride channel protein"
/protein_id="AAF00077.1"
/db_xref="GI:6002646"
/translation="MGSPRSSLFILVHLLEGAQNSLIQLNGNGVEGIVIAIDPNVP
EDERLIQIKDMVTFKASPLFEATEKRFYFNKVAIILPASWKAKEPYVKPKLEYKNA
DVVTEPNPENDGPTYEQMGCKEKEKIIYTPDFVAGKVLQYGPQGRVFEWHAH
LRWGVFNEYNEQKPYLSNKQPVICSAAIRGTNVLPQCQGGSCVTKPCRADRVETGL
FOKECEFIIPDOQSEKASIMFAQSIDTVVEFCCKEKNHKEAPNDQKNCRLNSTWEVI
ODESEFKTTPTTOPPAPTSLQIGQIVCLVDKSGSMTVGGRLKRLNOAGKLFEL
LOTVEGAWGVHAFDSAAAYVKSSELVQINSAAERDALARSLPTAASGGTSICGLRSA
FTVIRKKYPTDGEIIVLLTDGDNLTISACFPEVKQNGAIHTVALGPSAAKELEELUSQ
MTGGLOTYASDAQENGLIDAFGALSNGRAASQRSIQLESQGLTIQNNEMNGTVV
DSTVGKDTLITLERKFLSPFPFGVPSGRSQDSFLVGKHNKMAFYQVPGTAKVGMW
KYSLOASSOTLTLTVSSRRSSATLPDVTVTSMKMDTKGFPSPMVVYTKIHOCTLPIL
RAKVTALIESENGKTVTLELLDNGAGADATKNDGIYSRYETAVDANGYSVKVWALGG
VNTPRRAPPLWSGAMYIRGWIENGEIKNPPRPDINKDDLOQKQVCFRTASGGSFV
ASDVPSKPIPDLPFKCKITDLKAGIQGDNILNTWTAPGDDYDHGRADRYIIRISNI
LDLRKFNDSQVQNTDLIPKEANSEEVFKEGIPFTNGTDLFTAVQAVDKTNLKS
EISNIAQVSLFLPPEAPPETPPETAPSLPCPEIQVNSTIPGTHILKIMMKLGLQL
SIA"
BASE COUNT 892 a 758 c 755 g 674 t
ORIGIN

```

alignment_scores:

```

Quality: 3794.00 Length: 922
Ratio: 4.464 Gaps: 6
Percent Similarity: 92.191 Percent Identity: 78.850

```

alignment_block:

```

US-09-049-696-41 x AF095584 ..
Align seg 1/1 to: AF095584 from: 1 to: 3079
1 MetGlyProPheLysSerSerValPheIleLeuIleLeuHisLeuLeuGlu 17
|||||
133 ATGGGGTCATTAGAGTTCGCTGCTTCATCTGCTGCTTCACCTCTCGGA 182
|||||
17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34
|||||
183 AGGGGCTCAGAGTAACTCGCTCATCCAGCTGAATGGCAATGGCTACGAGG 232
|||||
34 lYlleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
|||||
233 GCATCGTCATTGCAATTGACCCCAACGTCGACAGAGATGAAGACTCAT 282
|||||
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
|||||
283 CAGAAATAAAGGATATGGTGACCAAAAGCATCTCCATATCTGTTTGAAGC 332
|||||
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGlu 84
|||||
333 TACGGAAAAAGATTCATTATTTCAAAAATGTTGCCATTTTGTATCTCTG 382
|||||
84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
|||||
383 GTTGGAGAGGCAAAACCGGATATGTGAACCAACCAACTTGAGACCTACAA 432
|||||
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
|||||
433 AATCCTCATGTTGGTGTACTGAACCAAAATCTCCAGAAATATGATGGCC 482
|||||
117 oTyrThrGluGlnMetGlyAsnCysGlyGluLysGluArgIleHisL 134
|||||

```

483 CTATACGGAGCAGATGGAACTCTGGAGAAAGGGTGAATAATTTATT 532
134 eutHrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
533 TCACCTCTGCTGCTCGTAGCAGGAAAGAGTCTCAATATGACCAAA 582
151 GlyArgAlaPheValHisGluTyrAlaHisLeuArgTrpGlyValPheAs 167
583 GGAAGGGTATTCTCCAGCAATGGGCTCACTCGCATGGGGCGTGTAA 632
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGlnA 184
633 TGAGTACAACAACAGCAGAGAAGTTCTATATATCCACAAAAAGACAAC 682
184 laValArgCysSerAlaGlyIleThrGlyThrAsnValValLysCys 200
683 CAGTGATATGTTCCGAGCAGCATCCGGGTACAAATGTACTACCACAATGC 732
201 GlnGlyCysSerCysTyrThrLysArgCysThrPheAsnLysValThrGI 217
733 CAAGAGGAGCAGTGTCTCACCACCGCTGCCGAGCAGACAGAGTAACGGG 782
217 yLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrGluL 234
783 GCTGTTCACAGAGGAGTGTACGTTCATATCCCGATCCCGAGCAGGAGA 832
234 ysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPheCys 250
833 AGGCCTCTATATGTTTGCACAGAGCATGTGACTGTGGTTGAATCTGT 882
251 ThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLysCy 267
883 AAGAAAAAACCACCAATTAAGAGCTCCAAACGACCAACCAAAATG 932
267 sAsnLysArgSerThrTrpGluValIleArgAspSerGluAspPheLysL 284
933 CAATCTCCGAAGACATGGGAAGTAGTATCAAGATTCTGAAGACTTTAAGA 982
284 ysThrThrProMetThrThrGlnProProAsnProThrPheSerLeuLeu 300
983 AAACCACTCCCATGACACACAGCCACCCGCGCCACCTTCTCAGTGTG 1032
301 GlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySerMe 317
1033 CAGATCGGACAGAGAAATCGTGTCTAGTCTCGACAAAGTCTGGAAGCAT 1082
317 tAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuPheL 334
1083 GACGGTTGGTGGCCGTCTTAAGCGACTGAATCAAGCGCAAACTCTTCC 1132
334 euLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPheAsp 350
1133 TTCTGCAGACAGTGGAGCAAGGGGCTGGGTGGATGGTGGCATTTGAC 1182
351 SerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySerAs 367
1183 AGCCTGCTATGTATTAAGTGAAGTCTGTCCAGATAAACACGCGCGCGGA 1232
367 pArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyThrS 384
1233 AAGAGACGCGCTTCCCGCAGCTTACCACAGCAGCCTCGGAGGAACGT 1282
384 erIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLysTyr 400
1283 CCATCTGCTCGGGGCTTCCGCTCCGCATTTACTGTGATCAAGAAATAAC 1332
401 ProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAsnTh 417
1333 CCGACCGAGCGGTCTGAGTCTGTACTGACCCAGCGGGAGGAGCAACAC 1382
417 rIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleHisT 434
1383 GATAAGCGCGTGTCTTCCGAGGAGTGAAGCAAAACGGGAGCCATCATCCA 1432

434 hrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluLeuSerLys 450
1433 CGGTGCCCTTGGCCCTCGGACGAAGAGAGCTGGAGGAGTGTGCGAG 1482
451 MetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAsnAsnGI 467
1483 ATGACAGGTGGTTTGCAGACATATGCTTCAGACCAGCGCGAGAACACGG 1532
467 yLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlaValSerG 484
1533 CCTCATGTGACGCTTTCGGGGCCCTTTCGTCGGGAACAGAGCTGCTCCC 1582
484 InArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsnSerGln 500
1583 AGCGCTCATCCAGCTGGAGAGTCAAGGCTTAACCTCCAGAACACAGAA 1632
501 TrpMetAsnGlyThrValIleValAspSerThrValGlyLysAspThrLe 517
1633 TGGATGAATGGCACAGTGTGTGGACAGCAGCTGTGGGAAAGACACTTT 1682
517 uPheLeuIleThrTrpThrThrGlnPro...ProGlnIleLeuLeuTr 532
1683 GTTCTCATACCTTGGACCGAAGTCTCTTCCCAATTCCTTTTGTG 1732
532 pAspProSerGlyGlnLysGlnGlyPheValValAspLysAsnThr 548
1733 GGGTTCACGCGGAAGACGAGCAGCTTCTGTGGTGGAAACACACAC 1782
549 LysMetAlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLy 565
1783 AAGATGGCTTACTTCCAAAGTCCAGGACGCGCAAGGTGGCATGTGAA 1832
565 sTyrSerLeuGlnAlaSerSerGlnThrLeuThrLeuThrValThrSera 582
1833 GTACAGCTGCACAGCAAGTTTCAAAACCTCCTTTGACCGTCACTGCC 1882
582 rgAlaSerAsnAlaThrLeuProProIleThrValThrSerLysThrAsn 598
1883 GTCCGTCCAGCGCTACCTGCTCGGTACCGTGACCTCTAAATGAAC 1932
599 LysAspThrSerLysPheProSerProLeuValTyrAlaAsnIleAr 615
1933 AAAGACAGCGGCAAAATCCCGAGCCCATGTTAGTTTACACGAGATGCA 1982
615 gGlnGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluS 632
1983 CCAAGAACCTGCCCCATTCAGGGCCCAAGGTACAGCCCTGATAGAT 2032
632 erValAsnGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGly 648
2033 CAGAAATGGGAAACAGATTACCTTGAATTTACTGGACACGAGCAGGC 2082
649 AlaAspAlaThrLysAspAspGlyValTyrSerArgTyrPheThrThrTy 665
2083 GCTCATGCTACCAAGAATGACGGTATCTACTCGAGGTATTTTACAGCTTA 2132
665 rAspThrAsnGlyArgTyrSerValLysValArgAlaLeuGlyGlyValA 682
2133 CGATGCAAAATGGTGCATACAGCGTAAAGTGTGGGCTCTGGAGGACTGA 2182
682 snAlaAlaArgArgArgValIleProGlnGlnSerGlyAlaLeuTyrIle 698
2183 ACAGCCACCCAGAGGCGACCCCTCTGGAGTGGAGCATGTACATA 2232
699 ProGlyTrpIleGluAsnAspGluIleGlnTrpAsnProProArgProGI 715
2233 CGTGGTGGATTGAGATGGGAGATCAAAATGGAACCCCGAGAGCCGA 2282
715 uIleAsnLysAspValGlnHisLysGlnValCysPheSerArgThrS 732
2283 CATTAACAGGAGCAGCTTCAGGGCAAGTGTGTGTTTCAGCCGAACAG 2332

117 oTyrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134
365 CTACCCGCAACATATATGAGAGCATGTGGAGAAAGGGATCAGGATTCACC 414
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrClyProGln 150
415 TGACTCTGACTCTTAGCAGGAAGAAGCTGACTCAGTATGGGCCACAA 464
151 GlyAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
465 GACAGGACCTTTGTCCATCAGTGGCTCACATCCGATGGGAGCTTTAA 514
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGlyArgIleGlnA 184
515 TGAATACAAACAGCAGCAAGACTTCTACTTATCCAAAGAAACCCCAAG 564
184 laValArgCysSerAlaGlyIleThrGlyThrAsnValValLysCys 200
565 CAGTGAGGTTCACGACGCATTACCGGTAAATCAAGTTCTGCTGGTGC 614
201 GlnGlyGlySerCysTyrTrpLys...ArgCysThrPheAsnLysValTh 216
615 CAGGAGGCAGTTGTATCACTAACGGAAGTGTGTAATCGACAGAGTAAC 664
216 rGlyLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrG 233
665 GGGACTGTATAAAGCAATTTGTATTGTACCAAGATCCACACCAAAACG 714
233 luLysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPhe 249
715 AGNAGCTTCATCATGTTTAACCAAAATATCAATCTGTGGTGTGAATTC 764
250 CysThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLy 266
765 TGTATCAAGAAAAATACAAATCAAGAAGCCCAATGACCAACCAACCAACG 814
266 sCysAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheL 288
815 ATGCAATCTCCGAAGCAGCTGGGAAGTCAATCCAGGAATCTCAGAGACTTCA 864
283 yLysThrThrProMetThrThrGlnProAsnProThrPheSerLeu 299
865 AGCAAACTCTCCATGATGACGCCAGCCACCCTGCACCCACCTCTCTACTG 914
300 LeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySe 316
915 CTGCAAAATGGCAAGAAATGTGTGCTTACTTCTGTAAGTCCGGGAG 964
316 rMetAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeuP 333
965 CATGCTCGACGATGATCGTCTTAACCGAATCAATCAGCAAGCCGGCTTT 1014
333 heLeuLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPhe 349
1015 TCTCTGCTGCACTCTGGAGCAGGATCCTGGGTGGGATGGTGAACCTTT 1064
350 AspSerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySe 366
1065 GACAGTGTGCTATGTACAAGCGAACTCAACACAGTAAACAGTGGTGC 1114
366 rAspArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyT 383
1115 TGACAGAGATCTGCTGATCAAGCACTTACCCACAGTATCTGCAGGAGGGA 1164
383 hrSerIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLys 399
1165 CATCTATATGCTTGGCCTTCGGACAGCATTTACAGTGTATTAAGAGAGAG 1214
400 TyrProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAspAs 416
1215 TATCCAACCTGATGATCTGAATTTGCTGCTGCTGACCATGGGAGGACAA 1264
416 nThrIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleIleH 433


```
2165 TAGTTAT.....GTTCAAGCAGCAGCTGCTGCTTCAGCAGGACATCTT 2208
733 erGlySerPheValAlaSerAspValPro...AsnAlaProIlePro 748
|||||.....
2209 CAGGGGATCGTTGTGGCCACCAATGTCGCCGAGCAGCTCCATTCCT 2258
749 AspLeuPheProGlyGlnIleThrAspLeuLysAlaGluIleHisG1 765
|||||.....
2259 GACCTCTTCCACCTCTCAATCACTGACCTGAAGCCAGCATCCAAGG 2308
765 yGlySerLeuIleAsnLeuThrTrpThrAlaProGlyAspAspThrAspH 782
|||||.....
2309 GCAGAACTCGTGNATCTGACGTGGAGCGCTCTGGGGATGACTACGACC 2358
782 isGlyThrAlaHisLysTyrlleIleArgleSerThrSerIleLeuAsp 798
|||||.....
2359 ACGGGAGAGCTTCAACTATCATCTCGAATGAGCAGCAGTATGCTGTGAT 2408
799 LeuArgAspLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuI1 815
|||||.....
2409 CTCAGGAGCAGCTTCAACACTCTCACTCCAAGTGAACACTACCGGTCTTAT 2458
815 eProLysGluAlaAsnSerGluGluValPheLeuPheLysProGluAsnI 832
|||||.....
2459 CCCAAGAGCCAGCTCTCAGGAAATCTTGAGTTGAACTGGGAGGCA 2508
832 leThrPheGluAsnGlyThrAspLeuPheIleAlaIleGlnAlaValAsp 848
|||||.....
2509 ACATTTTGGAAATGGCACAGATATCTTCATTTGCTATCCAGGCTGTGAT 2558
849 LysValAspLeuLysSerGluIleSerAsnIleAlaArgValSerLeuPh 865
|||||.....
2559 AAGTCCAATCTGAATCAGAAATCTCAACATTCGACGGGTGCTGTGTT 2608
865 eileProProGlnThrProProGluThrProSerProAspGluThrSerA 882
|||||.....
2609 CATCCCGCTCAGGAGCGG.....CCATTCGGAAGACTCAACTC 2649
882 laProCysProAsnIleHisIleAsnSerThrIleProGlyIleHisIle 898
|||||.....
2650 CCCTTTCTCTGATCAGCATCAACAGCACCATTCTCTGCATCCAGCTG 2699
899 LeuLysIleMetTrpLysTrpIleGlyGluLeuGlnLeuSerIle 913
|||||.....
2700 CTGAAGATAATGTGGAAGTGGCTAGGGGAAATGCAGGTGACACTA 2744
seq_name: gb_pr:AK024970
```

```
seq_documentation_block: 2022 bp mRNA PRI 29-SEP-2000
LOCUS AK024970
DEFINITION Homo sapiens cDNA: FLJ21317 fis, clone COL02275, highly similar to
AF127036 Homo sapiens calcium-activated chloride channel protein 1
(CaCC1) mRNA.
ACCESSION AK024970
VERSION AK024970.1 GI:10437397
KEYWORDS oligo capping: fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
REFERENCE Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
AUTHORS Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
TITLE Unpublished (2000)
REFERENCE 2 (bases 1 to 2022)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
TITLE Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
JOURNAL Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
```

Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT: NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Deparment of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES
.. source
.. 1..2022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL02275"
/clone_lib="COL"
/tissue_type="colon"
/note="cloning vector pME18SFL3"
misc_feature
1..2022
/note="highly similar to AF127036 Homo sapiens
calcium-activated chloride channel protein 1 (CaCC1) mRNA"
BASE COUNT 612 a 472 c 453 g 485 t
ORIGIN

alignment_scores:
Quality: 3357.00 Length: 651
Ratio: 5.157 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-41 x AK024970 ..
Align_seg 1/1 to: AK024970 from: 1 to: 2022
264 AsnGlnLysCysAsnLeuArgSerThrTrpGluValIleArgAspSerG1 280
|||||.....
3 ATCAAAATGCAATCTCCGAAGCAGCATGGAGTGGATCGGTGATCTGA 52
280 uAspPheLysLysThrThrProMetThrThrGlnProProAsnProThrP 297
|||||.....
53 GGACTTTAAGAAAACCACTCTCTATGACAACACAGCCACCAATCCACCT 102
297 heSerLeuLeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLys 313
|||||.....
163 TCTCATTTGCTGCAGATTGGACAAAAGAAATTTGTTAGTCTCTTGACAAA 152
314 SerGlySerMetAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaG1 330
|||||.....
153 TCTGGAAGCATGGCGACTGGTAACCGCTCAATCGACTGAATCAAGCAGG 202
330 yGlnLeuPheLeuLeuGlnThrValGluLeuGlySerTrpValGlyMetV 347
|||||.....
203 CCAGCTTTCTCTGCTGCAGACAGTTGAGTGGGCTCTCGGTGGATGG 252
347 alThrPheAspSerAlaAlaHisValGlnSerGluLeuIleGlnIleAsn 363
|||||.....
253 TGACATTTGACAGTCTGCCCATGTACAAAGTGAACATCAGATATAAC 302
364 SerGlySerAspArgAspThrLeuAlaLysArgLeuProAlaAlaLase 380
|||||.....
303 AGTGGCAGTGACAGGGACACACTGCCAAAAGATTACTCGCAGCAGCTTC 352
380 rGlyGlyThrSerIleCysSerGlyLeuArgSerAlaPheThrValIleA 397
|||||.....
353 AGGAGGACGCTCCATCTGCAGCGGCTTCGATCGCATTTACTGTGATTA 402
397 rgLysLysTyrrProThrAspGlySerGluIleValLeuLeuThrAspGly 413
|||||.....
403 GGAAGAAATATCCAACTGATGGATCTGAAATTTGCTGCTGACGATGGG 452
414 GluAspAsnThrIleSerGlyCysPheAsnGluValLysGlnSerGlyAl 430
|||||.....

453 GAAGACACACTATAAGTGGTGCTTTAAACGAGGTCAAAACAAAGTGGTCC 502
430 aillelleHlstrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluG 447
503 CATCATCCACACAGTCGGCTTTGGGGCCCTCGCAGCTCAAGAACTAGAGG 552
447 luleuSerLysMetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnVal 463
553 AGCTGTCCAAATGACAGGAGGTTTACAGACATATGCTTCAGATCAAGTT 602
464 GlnAsnAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnG 480
603 CAGAAATGGCTCATTCATGCTTTTGGGGCCCTTTTCATCAGGAATGG 652
480 yAlaValSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuG 497
653 AGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAAGGATTAACCCCTCC 702
497 lInAsnSerGlnTrpMetAsnGlyThrValIleValAspSerThrValGly 513
703 AGACACGCGAGTGGATGATGACAGCTGATCGTGACAGCAGCCGTGGGA 752
514 LysAspThrLeuPheLeuIleThrTrpThrGlnProGlnIleLeu 530
753 AAGGACACTTGTCTTATCATCCTGGACACGACGAGCTCCCCAAATCCT 802
530 uLeuTrpAspProSerGlyGlnLysGlnGlyPheValValAspLysA 547
803 TCTCTGGGATCCAGTGGACAGAACAGTGGCTTTGTAGTGGACAAA 852
547 snThrLysMetAlaTrpLeuGlnIleProGlyIleAlaLysValGlyThr 563
853 ACACCAAAATGGCTTACCTCCAAATCCCAAGGCAATGCTTAAGTTGGCACT 902
564 TrpLysTrpSerLeuGlnAlaSerSerGlnThrLeuThrLeuThrValTh 580
903 TGGAAATACAGTCTCAAGCAAGCTCACAAACCTTGACCCCTGACTGTAC 952
580 rSerArgAlaSerAsnAlaThrLeuProProIleThrValThrSerLysT 597
953 GTCCCGTGGCTCAATGTACCTCCCTCCCAATTACAGTGACTTCCAAA 1002
597 hrAsnLysAspThrSerLysPheProSerProLeuValValTrpAlaAsn 613
1003 CGACACAGGACACAGCAAAATCCCAAGCCCTCTGTGTATGTAAT 1052
614 lIleArgGlnGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuI 630
1053 ATTCGCCAAGGAGCCTCCCAATTCACAGGCCAGTGTACAGCCCTGAT 1102
630 egluSerValAsnGlyLysThrValThrLeuGluLeuLeuAspAsnGlyA 647
1103 TGAATCAGTGAATGAAACAGTACCTTGGAACTACTGGATAATGGAG 1152
647 laGlyAlaAspAlaThrLysAspAspGlyValTyrSerArgTyrPheThr 663
1153 CAGGTGCTGATCTACTAAGGATGAGGTGTCTACTCAAGGTATTTTACA 1202
664 ThrTyrAspThrAsnGlyArgTyrSerValLysValArgAlaLeuGlyG 680
1203 ACTTATGACACGAATGAGATACAGTGTAAAGTGCAGGCTCTGGAGG 1252
680 yValAsnAlaAlaArgArgValIleProGlnGlnSerGlyAlaLeuT 697
1253 AGTTAACCAGCCAGCAGGAGTGTATCCCGCAGAGTGGAGCACTGT 1302
697 yrIleProGlyTrpIleGluAsnAspGluIleGlnTrpAsnProProArg 713
1303 ACATACCTGGTGGATTGAGATGATGAATACATGGAATGCCACCAAGA 1352
714 ProGluIleAsnLysAspValGlnHisLysGlnValCysPheSerAr 730
1353 CCTGAAATTAATAGGATGTGTTCAACACAGCAAGTGTGTTTCAGCAG 1402

730 gThrSerSerGlyGlySerPheValAlaSerAspValProAsnAlaProI 747
1403 AACATCTCGGAGGCTCATTTGTGCTTCTGATGTCCCAAAATGCTCCCA 1452
747 leProAspLeuPheProGlyGlnIleThrAspLeuLysAlaGluIle 763
1453 TACCTGATCTCTCCACCTGGCCAAATCACCGACCTGAAGCGGGAATTT 1502
764 HisGlyGlySerLeuIleAsnLeuThrTrpThrAlaProGlyAspAspTy 780
1503 CACGGGGGCGAGTCTCATTAATCTGACTTGGACAGCTCCTGGGGATGATTA 1552
780 rAspHisGlyThrAlaHisLysTyrIleIleArgIleSerThrSerIleL 797
1553 TGACCATGGACAGCTCACAGTATATCATTCGAATAAGTACAAGTATTC 1602
797 euAspLeuArgAspLysPheAsnGluSerLeuGlnValAsnThrThrAla 813
1603 TTGATCTCAGAGACAAGTTCAATGAATCTTCAAGTGAATACTACCGCT 1652
814 LeuIleProLysGluAlaAsnSerGluGluValPheLeuPheLysProG 830
1653 CTCATCCCAAGGAAGCCAACTCTGAGGAAGTCTTTTCTTTTAAACCAGA 1702
830 uAsnIleThrPheGluAsnGlyThrAspLeuPheIleAlaIleGlnAlav 847
1703 AAACATTAATCTTTGAAAATGGCACAGATCTTTTCATTGCTATTCAGGCTG 1752
847 aIAspLysValAspLeuLysSerGluIleSerAsnIleAlaArgValSer 863
1753 TTGATAAGGTGCTGATCTGAAATCAGAAATATCAACATTCACGAGTATCT 1802
864 LeuPheIleProProGlnThrProProGluThrProSerProAspGluTh 880
1803 TTGTTTATTCCTCCACAGACTCCGCCAGACAGACACCTAGTCTGTAGAAC 1852
880 rSerAlaProCysProAsnIleHisIleAsnSerThrIleProGlyIleH 897
1853 GTCTGCTCTTGTCTTAATATTCATATCATATCAACAGCACCATTCTCGCATTC 1902
897 lsIleLeuLysIleMetTrpLysTyrIleGlyGluLeuGlnLeuSerIle 913
1903 ACATTTTAAAAATTTATGGAAGTGGATAGGAGAACTGCAGCTGTCAATA 1952
914 Ala 914
1953 GCC 1955

seq_name: gb_pat:AX092338

seq_documentation_block:
LOCUS AX092338 3265 bp DNA PAT 21-MAR-2001
DEFINITION Sequence 69 from Patent WO0116318.
ACCESSION AX092338
VERSION AX092338.1 GI:13444483
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Eaton,D.L., Filvaroff,E., Gerritsen,M.E., Goddard,A.,
Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and
Wood,W.I.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0116318-A 69 08-MAR-2001;
Genentech, Inc. (US)
FEATURES
source Location/Qualifiers
1..3265
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 1159 a 596 c 632 g 878 t
ORIGIN

alignment_scores:

Quality: 2879.50 Length: 914
Ratio: 3.779 Gaps: 8
Percent Similarity: 83.370 Percent Identity: 61.597

alignment_block:

US-09-049-696-41 x AX092338 ..

Align seg 1/1 to: AX092338 from: 1 to: 3265

```
1 MetGlyProPheLysSerValPheLeuLeuLeuLeuLeuLeuL 17
25 ATGGGGTATTTCAGAGTCTTTGTTTCTCTAGTCTGTGCTGCTGCA 74
17 uGlyAlaLeuSerAsnSerLeuLeuLeuLeuLeuLeuLeuLeu 34
: : : : : : : : : : : : : : : : : : : : : : : : :
75 CCAGTCA...ANTACTTCCTTCATAGCTGAATAATAATGGCTTTGAAG 121
34 lyileValValAlaileAspProAsnValProGluAspGluThrLeu 50
: : : : : : : : : : : : : : : : : : : : : : : : :
122 ATATTGTCATTGTTATAGATCTAGTGTGCCAGAGATGAAAAATAATT 171
51 GlnGlnLeuLysAspMetValThrGlnAlaSerLeuTyrLeuPheGlu 67
: : : : : : : : : : : : : : : : : : : : : : : : :
172 GAACAAATAGAGGATATGGTGATACAGCTTCTACGTACCTGTTTGAAGC 221
67 aThrGlyLysArgPheTyrPheLysAsnValAlaLeuLeuLeuProGlu 84
: : : : : : : : : : : : : : : : : : : : : : : : :
222 CACAGAAAAAGATTTTTTCAAAATGATCTATATTAATTCCTGAGA 271
84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
: : : : : : : : : : : : : : : : : : : : : : : : :
272 ATTGGAAGGAAATCTCAGTACACAAAGGCCAAACATGAAACCATAAA 321
101 AsnAlaAspValLeuValAlaGluSerThrProProGlyAsnAspGluPr 117
: : : : : : : : : : : : : : : : : : : : : : : : :
322 CATGCTGATGTTATAGTTGACACCTACCTCCAGGTAGATGAACCC 371
117 oTyrThrGluGlnMetGlyAsnCysGlyGlyLysGlyGluArgileHisL 134
: : : : : : : : : : : : : : : : : : : : : : : : :
372 ATACACCAAGCAGTTCACAGATGTGAGAGAAAGGCGAATACATTCACT 421
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
: : : : : : : : : : : : : : : : : : : : : : : : :
422 TCACCCCTGACCTTCTACTTGCAGAAAAAATAATGAATATGGACCA 471
151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
: : : : : : : : : : : : : : : : : : : : : : : : :
472 GGCAAACTGTTTGTCCATAGTGGGCTCACCTCCGGTGGGAGTGTGA 521
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGly...ArgileG 183
: : : : : : : : : : : : : : : : : : : : : : : : :
522 TGAGTACAAATGAATCAGCCCTTCTACCGTGCTAAGTCAAAAAATCG 571
183 lnAlaValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLys 199
: : : : : : : : : : : : : : : : : : : : : : : : :
572 AAGCAACAAGGTGTTCCGACGATATCTGTGTGAGAAATAGAGTTTATAAG 621
200 CysGlnGlyGlySerCysTyrThrLysArgCysThrPheAsnLysValTh 216
: : : : : : : : : : : : : : : : : : : : : : : : :
622 TGTCAGGAGGACGCTGCTCTAGTAGAGCATGCAGAAATTTGATTACAAC 671
216 rGlyLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrG 233
: : : : : : : : : : : : : : : : : : : : : : : : :
672 AAAACTGTATGGAAAGATGTTCAATTTCTTTCTGATAAAGTACAAACAG 721
233 luLysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPhe 249
: : : : : : : : : : : : : : : : : : : : : : : : :
722 AAAAGGATCCATAATGTTATGCAAGATATTGATTCTGTGTTGAATTT 771
```

```
250 CysThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLy 266
: : : : : : : : : : : : : : : : : : : : : : : : :
722 TGTAAAGAAAAAACCCATAATCAAGAGCTCAAGCCTACAAAACATAAA 821
266 SCysAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheL 283
: : : : : : : : : : : : : : : : : : : : : : : : :
822 GTGCAATTTTACAGATACATGGGAGGTATTAGCAATTTCTGAGGATTTTA 871
283 ySLysThrThrProMetThrThrGlnProProAsnProThrPheSerLeu 299
: : : : : : : : : : : : : : : : : : : : : : : : :
872 AAAACACATACCCATGGTGACACCACTCCTCCACCTGTCTCTCATTTG 921
300 LeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySe 316
: : : : : : : : : : : : : : : : : : : : : : : : :
922 CTGAAGATCAGTCAAGAAATTTGTGTGTAGTTCTTGTATAAGTCTGGAAG 971
316 rMetAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyGlnLeup 333
: : : : : : : : : : : : : : : : : : : : : : : : :
972 CATGGGGGTAAAGACCCCTTAATCGAATCGAATCAAGCAGCAACACATT 1021
333 heLeuLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPhe 349
: : : : : : : : : : : : : : : : : : : : : : : : :
1022 TCCTGCTGCAGACGTGTAATAATGATCTCTGGTGGGATGTTTCACATT 1071
350 AspSerAlaAlaHisValGlnSerGluLeuLeuGlnIleAsnSerGlySe 366
: : : : : : : : : : : : : : : : : : : : : : : : :
1072 GATAGTACTGCCACTATTGTAATAATGATCAATCCAAATAAAAGCAGTGA 1121
366 rAspArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyT 383
: : : : : : : : : : : : : : : : : : : : : : : : :
1122 TGAAGAAACACACTCATGCCAGGATTACCTACATATCTCTGGGAGGAA 1171
383 hrSerIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLys 399
: : : : : : : : : : : : : : : : : : : : : : : : :
1172 CTTCATCTCTCTCGAATTAATATGATTCAGTTAGTTGGAGAGCTA 1221
400 Tyr...ProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAs 415
: : : : : : : : : : : : : : : : : : : : : : : : :
1222 CATTCCCAACTCGATGGATCGAAGTACTGCTGCTGATGGGAGGA 1271
415 pAsnThrIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleI 432
: : : : : : : : : : : : : : : : : : : : : : : : :
1272 TAACACTCCAAGTCTTGTATTGATGAAGTCAACAAAGTGGGCCCATG 1321
432 leHisThrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluLeu 448
: : : : : : : : : : : : : : : : : : : : : : : : :
1322 TTCATTTTATTGCTTTGGGAAGAGCTGCTGATGAAGCAGTAATAGAGATG 1371
449 SerLysMetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAs 465
: : : : : : : : : : : : : : : : : : : : : : : : :
1372 AGCAAGATAACAGAGGAAAGTCAATTTTATGTTTCAGATGAAGCTCAGAA 1421
465 nAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlav 482
: : : : : : : : : : : : : : : : : : : : : : : : :
1422 CAATGGCCTCATTCATGCTTTTGGGGCTCTTACATCAGGAATACTGATC 1471
482 alSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsn 498
: : : : : : : : : : : : : : : : : : : : : : : : :
1472 TCTCCACAGAGTCCCTTCAGCTCGAAAGTAAGGATTAAACACTGAATAGT 1521
498 SerGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLysAs 515
: : : : : : : : : : : : : : : : : : : : : : : : :
1522 AATCCCTGGATGAACGACACTGTCATAATTTAGTAGTACAGTGGGAAAGGA 1571
515 pThrLeuPheLeuIleThrTrpThrGlnProProGlnIleLeuLeut 532
: : : : : : : : : : : : : : : : : : : : : : : : :
1572 CACGTTCTTCTCATCATCATGGAACAGTCTGCCCTCCAGCATTTCTCTCT 1621
532 rpAspProSerGlyGlnLysGlnGlyPheValValAspLysAsnThr 548
: : : : : : : : : : : : : : : : : : : : : : : : :
1622 GGGATCCACGTGGAACAATAATGGAATAATTCACAGTGGATGCAACTTCC 1671
```

549 LysMetAlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLy 565
|||||
1672 AAATGGCTATCTCAGTATTCAGGAAGTCAAAAGGTGGCACTGGGC 1721
565 sTyrSerLeuGlnAlaSerSer.....GlnThrLeuThrValTh 580
|||||
1722 ATACAATCTTCAAGCCAAAGGCAACCCAGAAACATTAACTATTACAGTAA 1771
580 hrSerArgAlaSerAsnAlaThrLeuProPheIleThrValThrSerLys 596
|||||
1772 CTTCTGAGCAGCAAAATCTCTCTGCTGCTCAATCAAGTGAATGCTAAA 1821
597 ThrAsnLysAspThrSerLysPheProSerProLeuValValThrAlaAs 613
|||||
1822 ATGAATAGACGAGTAAACAGTTTCCCGAGCCCAATGATTGTTACGCAGA 1871
613 nIleArgGlnGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuI 630
|||||
1872 AATTCTACAAGGATATGATCTGTCTTGGAGCCCAATGTCACCTTTTCA 1921
630 leGluSerValAsnGlyThrValThrValThrLeuGluLeuLeuAsnGly 646
|||||
1922 TTGATCACAGATGGACATACAGAGTTTGGAACTTTTGGATAATGGT 1971
647 AlaGlyAlaAspAlaThrLysAspGlyValThrValThrValThrPheTh 663
|||||
1972 CGAGGCGCTATCTTCAAGATGATGGAGTCTACTCCAGGTATTTTAC 2021
663 rThrTyrAspThrAsnGlyArgTyrSerValLysValArgAlaLeuGlyG 680
|||||
2022 AGCATATACAGAAATGGCAGATATAGCTTAAAGTTGGGCTCATGGAG 2071
680 lyValAsnAlaAlaArgArgValIleProGlnGlnSerGlyAlaLeu 696
|||||
2072 GAGCAACACTGCAGCTGAAATAGGCTCCCTCACTGAATAGAGCGCG 2121
697 TyrIleProGlyTrpIleGluAsnAspGluIleGlnTrpAsnProAr 713
|||||
2122 TACATACAGCGTGGTGTAGTGAACGGGAAATGAAGCAACCCGCCAG 2171
713 qProGluIleAsnLysAspValGlnHisLysGlnValCysPheSerA 730
|||||
2172 ACTGAAATTTGAT...GAGGACTACTCAGCACCACTTGGAGGATTTTCAGCC 2218
730 rgThrSerGlyGlySerPheValAlaSerAspValProAsnAlaPro 746
|||||
2219 GAACAGCATCGGAGGTGATTTGTGTATCACAAGTCCCAAGCTTCCC 2268
747 IleProAspLeuPheProGlyGlnIleThrAspLeuLysAlaGluI 763
|||||
2269 TTGCCTGACCAATACCCCAAGTCAAAATCAGACACCTTGATGCCACAGT 2318
763 eHisGlyGlySerLeuIleAsnLeuThrTrpThrAlaProGlyAspAs 780
|||||
2319 TCATAGGATAGATATT...CTTACATGGACAGCAGCAGGAGGATATT 2365
780 yAspHisGlyThrAlaHisLysTyrIleIleArgIleSerThrSerIle 796
|||||
2366 TTGATGTTGGAAAGTTCAACGTTATATCATAGAATAAGTCAAGTATT 2415
797 LeuAspLeuArgAspLysPheAsnGluSerLeuGlnValAsnThrAla 813
|||||
2416 CTGTATCTAGACAGATTTTGTATGCTCTTCAAGTAAATACTACTGA 2465
813 aLeuIleProLysGluAlaAsnSerGluGluValPheLeuPheLysPro 830
|||||
2466 TCTGTCAACAAAGGAGGCAACTCAAGAAAGCTTTCATTTAAACCCAG 2515
830 luAsnIleThrPheGluAsnGlyThrAspLeuPheIleAlaIleGlnAla 846
|||||
2516 AAAATATCTCAGAGAAATGCAACCCACATATTTATGCCATTAAAAAGT 2565
847 ValAspLysValAspLeuLysSerGluIleSerAsnIleAlaArgValSe 863

2566 ATAGATAAAAGCAATTGACATCAAAAGTATCCAAACATTGCCAAGTAAC 2615
863 rLeuPheIleProGlnThrProGlu...ThrProSerProAspG 879
|||||
2616 TTTGTTTATCCCAAGCAAACTCTGATGACATTGATCTTACACCTACTC 2665
879 luthrSerAlaProCysProAsnIleHisIleAsnSerThrIleProGly 895
|||||
2666 CTACTCTACTCTACTCTCTGATAAAGTCATAATCT.....GGA 2706
896 IleHisIleLeuLysIleMetThrLysTrpIleGlyGluLeu 909
|||||
2707 GTTAATATTCTACCGTGGTATTGTTCTGTCATTGGGTCGTCT 2748

seq_name: gb_pr:AK000072

seq_documentation_block: 3221 bp mRNA PRI 22-FEB-2000
LOCUS AK000072 Homo sapiens cDNA FLJ20065 fis, clone COL01613, highly similar to
DEFINITION EC1C_BOVIN EPITHELIAL CHLORIDE CHANNEL PROTEIN.

ACCESSION AK000072
VERSION AK000072.1 GI:7019922
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL01613.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (sites)
AUTHORS Kawabata,A., Hikiiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 3221)

AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2000) to the DDBJ/EMBL/Genbank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo, Department
of Virology, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)

COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES Location/Qualifiers

source 1..3221
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL01613"
/clone_lib="COL"
/tissue_type="colon"
/note="cloning vector pME18SFL3"

misc_feature 1..3221

/note="highly similar to EC1C_BOVIN EPITHELIAL CHLORIDE
CHANNEL PROTEIN"

BASE COUNT 1105 a 600 c 634 g 882 t
ORIGIN

alignment_scores:

Quality: 2877.50 Length: 914
Ratio: 3.781 Gaps: 8
Percent Similarity: 83.260 Percent Identity: 61.597

alignment_block:

US-09-049-696-41 x AK000072

Align seg 1/1 to: AK000072 from: 1 to: 3221

1 MetGlyProPheLysSerSerValPheIleLeuIleLeuHisLeuLeuG1 17
||||| : : : : : ||||| : : : : : ||||| : : : : :
43 ATGGGGTATTTCAGAGGTTTGTCTTCTTCTTCTTCTTCTGCTGCTGCA 92
17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGlUG 34
:
93 CCAGTCA...AATACTTCCCTCATAGCTGAATAAATAGGCTTTGAAG 139
34 lyIleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
||||| : : : : : ||||| : : : : : ||||| : : : : :
140 ATATTGTCATTGTTATAGATCTAGTGTGCCAAGATGAAAAATAATT 189
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
:
190 GAACAAATAGAGGATATGGTACCTACAGCTTCTACGTACCTGTTTGAAGC 239
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluT 84
||||| : : : : : ||||| : : : : : ||||| : : : : :
240 CACAGAAAAAGATTTTTTCAAAATGATCTATATTAATCTCTGAGA 289
84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
:
290 ATTGCAAGGAAATCTTCAGTACAAAAGGCCAAACATGAAACCAATAAA 339
101 AsnAlaaspValLeuValAlaGluSerThrProProGlyAsnAspGluPr 117
:
340 CATGCTGATGTTATAGTGTGCCACCACTACATCTCCAGGTAGATGAACC 389
117 oTyrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134
||||| : : : : : ||||| : : : : : ||||| : : : : :
390 ATACCCACAGCAGTTCACAGATGTGGAGAAAGGCGAATACATCACT 439
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
:
440 TCACCCCTGACCTTCTACTTGGAAAAAACAATAATGAATATGGACCACA 489
151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
:
490 GGCAAGCTGTTTGTCCATGATGTGGGCTCACCTCCGGTGGGAGTGTGTA 539
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGly...ArgIleG 183
||||| : : : : : ||||| : : : : : ||||| : : : : :
540 TGAGTACAAATGAAGATCAGCCTTCTACCGTGTAACTCAAAAAAATCG 589
183 InAlaValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLys 199
:
590 AAGCAACAAGGTGTTCCGAGGTATCTCTGTAGAAATAGAGTTTATAAG 639
200 CysGlnGlyGlySerCysTyrThrLysArgCysThrPheAsnLysValTh 216
||||| : : : : : ||||| : : : : : ||||| : : : : :
640 TGTCAAGGAGCGACTGTCTTAGTAGAGCATGCGAGAATTGATTTACAAC 689
216 rGlyLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrG 233
||||| : : : : : ||||| : : : : : ||||| : : : : :
690 AAAACTGTATGGAAGATGTCAATTCTTCTCGATAAAGTACAAACAG 739
233 luLysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPhe 249
:
740 AAAAGCATCCATAATGTTTATGCAAGATATTGATTCTGTGTTGAATT 789
250 CysThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLy 266
:
790 TGTAAACGAAAAAACCATAATCAAGAGCTCCAAAGCCTACAAAAACATAA 839
266 sCysAsnLeuArgSerThrTrpGluValIleArgAspSerClnAspPheL 283
||||| : : : : : ||||| : : : : : ||||| : : : : :
840 GTGCAATTTTAGAAGTACATGGGAGGTGATTAGCAATTCAGAGATTTA 889
283 ysLysThrThrProMetThrGlnProProAsnProThrPheSerLeu 299
:
890 AAAACACCATACCATGGTGACACCACTCTCCACCTGCTCTCATTTG 939

309 LeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlySe 316
||||| : : : : : ||||| : : : : : ||||| : : : : :
940 CTGAAGATCCGCTCAAGAATTGTGTCTAGTTCTTCTGATAAGTCTGAAG 989
316 rMetAlaThrClyAsnArgLeuAsnArgLeuAsnGlnAlaClyGlnLeup 333
:
990 CATGGGGGTAAAGACCCCTTAAATCAATGAATCAAGCAGCAAAACATT 1039
333 heLeuLeuGlnThrValGluLeuGlySerTrpValGlyMetValThrPhe 349
||||| : : : : : ||||| : : : : : ||||| : : : : :
1040 TCCTGCTGCACAGCTGTGAAATGATCCTGGTGGGATGTTCACTTT 1089
350 AspSerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSerGlySe 366
:
1090 GATAGTACTGCCACTATTGTAATTAAGCTAATCCAAATAAAAGCAGTGA 1139
366 rAspArgAspThrLeuAlaLysArgLeuProAlaAlaSerGlyGlyT 383
:
1140 TGAAGAAACACACTCATGCCAGGATTACCTACATATCTCTCGGAGAA 1189
383 hrSerIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLys 399
||||| : : : : : ||||| : : : : : ||||| : : : : :
1190 CTTCCATCTGCTCTGGAATTAAATATCATTTCAAGTTCAGTGGAGAGCTA 1239
400 Tyr...ProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAs 415
:
1240 CATTCCCAACCTCGATGATCCGAAGTACTGCTGCTGACTGATGGGAGGA 1289
415 pAsnThrIleSerGlyCysPheAsnGluValLysGlnSerGlyAlaIleI 432
||||| : : : : : ||||| : : : : : ||||| : : : : :
1290 TAACACTGCAAGTCTTGTATTGATGAAGTGAACAAAGATGGGCCCATG 1339
432 leHisThrValAlaLeuGlyProSerAlaAlaGlnGluLeuGluLeu 448
:
1340 TTCATTTTATTTGCTTTGGGAAGAGCTGCTGATGAAGCAGTAATAGAGATG 1389
449 SerLysMetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAs 465
:
1390 AGCAAGATAACAGGAGGAAGTCATTTTATGTTTCAGATGAAGCTCAGAA 1439
465 nAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlav 482
||||| : : : : : ||||| : : : : : ||||| : : : : :
1440 CAATGGCCTCATGTGCTTTTGGGGCTCTTACATCAGGAAATACTGATC 1489
482 alSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsn 498
:
1490 TCTCCAGAAAGTCCCTTCAGCTCGAAAGTAAGGATTAACACTGAATAGT 1539
499 SerGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLysAs 515
:
1540 AATGCCTGGATGAACGACACTGTCTAATTGATAGTACAGTGGGAGGA 1589
515 pThrLeuPheLeuIleThrTrpThrGlnProProGlnIleLeuLeu 532
||||| : : : : : ||||| : : : : : ||||| : : : : :
1590 CACGTTCTTCTCATCATCAGAACAGTCTGCCCTCCAGCATTTCTCTCT 1639
532 rpAspProSerGlyGlnLysGlnGlyPheValValAspLysAsnThr 548
:
1640 GGGATCCAGTGGAAACAATAATGAAAAATTTACAGTGGATGCAACTTCC 1689
549 LysMetAlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLy 565
:
1690 AAAATGGCTATCTCAGTATTCCAGGAACATGCAAGGTGGCAGCTTGGC 1739
565 sTyrSerLeuGlnAlaSerSer...GlnThrLeuThrLeuThrVal 580
||||| : : : : : ||||| : : : : : ||||| : : : : :
1740 ATACAATCTTCAAGCCAAAGCGAACCCAGAAACATTAACATTACAGTAA 1789
580 hrSerArgAlaSerAsnAlaThrLeuProProIleThrValThrSerLys 596
||||| : : : : : ||||| : : : : : ||||| : : : : :
1790 CTTCTCAGCAGCAAAATCTTCTCTGCTCCCAATCAGATCAATGCTAAA 1839

```
597 ThrAsnLysAspThrSerLysPheProSerProLeuValValThrAlaAs 613
|||||
1840 ATGATTAAGACGCTAAGACATCTCCCGAGCCCAATGATTGTTACGACGA 1889

613 nileArgGlnGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuI 630
|||||
1890 AATTCTACAAGGATATGCTACCTGTTTGGAGCCCAATGTGACATGCTTCA 1939

630 leGluSerValAsnGlyLysThrValThrLeuGluLeuLeuAspAsnGly 646
|||||
1940 TTGAATCACAGATGACATACAGAAAGTTTGGAACTTTTGGATAATGTT 1989

647 AlaGlyAlaAspAlaThrLysAspGlyValThrValThrSerArgThrPhe 663
|||||
1990 GCAGGCGCTGATCTTTCAGAAATGATGAGTCTACTCCAGGATTTTAC 2039

663 rThrTyAspThrAsnGlyArgTyrSerValLysValArgAlaLeuGlyG 680
|||||
2040 AGCATATACAGAAATGGCAGATATAGCTTAAAGTTTCGGGCTCATGGAG 2089

680 lyValAsnAlaAlaArgArgValIleProGlnGlnSerGlyAlaLeu 696
|||||
2090 GAGCAACACTGCGAGGCTAAATACGGCTCCACTGAATAGAGCGCG 2139

697 TyrIleProGlyTrpIleGluAsnAspGlyIleGlnTrpAsnProAr 713
|||||
2140 TACATACAGGCTGGTAGTGAACGGGGAATGAGCAACCCGCCAAG 2189

713 gProGluIleAsnLysAspValGlnHisLysGlnValCysPheSerA 730
|||||
2190 ACCTGAAATGAT...GAGGATACCTCAGACCACTTGGAGGATTTTCAGCC 2236

730 rGThrSerSerGlyLysPheValAlaSerAspValProAsnAlaPro 746
|||||
2237 GAACAGATCCGGAGTGCTTTGGTATCACAGTCCCAAGCCCTCC 2286

747 IleProAspLeuPheProGlyGlnIleThrAspLeuLysAlaGluI 763
|||||
2287 TTGCTGTGACCAATACCCCAAGTCAATACAGACCTTGTGACACAGT 2336

763 eHisGlyGlySerLeuIleAsnLeuThrTrpThrAlaProGlyAspAs 780
|||||
2337 TCATGAGGATAGATTAT...CTTACATGGACAGCAGCAGAGATAAT 2383

780 yAspHisGlyThrAlaHisLysTyrIleArgIleSerThrSerIle 796
|||||
2384 TTGATGTTGAAAAGTCAACGCTTATATCATAGAAATAGTCAAGTAT 2433

797 LeuAspLeuArgAspLysPheAsnGluSerLeuGlnValAsnThrAla 813
|||||
2434 CTGTATCTAAGACAGATTTTGTATGATGCTCTCAAGTAAATACTACTGA 2483

813 aLeuIleProLysGluAlaAsnSerGluGluValPheLeuPheLysPro 830
|||||
2484 TCTGTGTCACCAAGAGGAGCACTCCAGAAAGCTTTCGATTTAAACCA 2533

830 luAsnIleThrPheGluAsnGlyThrAspLeuPheIleAlaIleGlnAla 846
|||||
2534 AAAATATCTCAGAAAATGCAACCCACATATTTATGCCATTAAAGT 2583

847 ValAspLysValAspLeuLysSerGluIleSerAsnIleAlaArgValSe 863
|||||
2584 ATAGATAAAGCAATTTGACATCAAAAGATATCCAACTATGCACAAAGTAA 2633

863 rLeuPheIleProGlnTrpProGlu...ThrProSerProAspG 879
|||||
2634 TTTGTTTATCCCTCAGCAATCTGATGACATTCATCTCCTACCTACTC 2683

879 luThrSerAlaProCysProAsnIleHisIleAsnSerThrIleProGly 895
|||||
2684 CTACTCTACTCTACTCTCATAAAGTCAATAATCT...GGA 2724

896 lleHisIleLeuIleMetTrpLysTrpIleGlyGluLeu 909
```

```
seq. name: gb_pr:AF127035
2725 GTTAAATATTCTACGCTGATTGCTGCTGATTGGCTGCTGTT 2766
|||||
|||||
```

seq. documentation_block: 3204 bp mRNA PRI 11-AUG-1999

LOCUS AF127035 Homo sapiens calcium-activated chloride channel protein 2 (CaCC2)
DEFINITION mRNA, complete cds.

ACCESSION AF127035
VERSION AF127035.1 GI:5726288

KEYWORDS human.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 3204)

Agnel, M., Vermaat, F., and Culouscou, J.-M.

Identification of three novel members of the calcium-dependent
chloride channel (CaCC) family predominantly expressed in the
digestive tract and trachea

FEBS Lett. 455 (3), 295-301 (1999)

PUBMED 99364503

10437792

REFERENCE 2 (bases 1 to 3204)

Agnel, M. and Culouscou, J.-M.

Direct Submission

Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des
Carrieres, Rueil-Malmaison 92500, France

LOCATION/Qualifiers

1..3204

/organism="Homo sapiens"

/db_xref="taxon:9606"

/tissue_type="colon"

1..3204

/gene="CaCC2"

29..2782

/gene="CaCC2"

/note="bovine epithelial chloride channel homolog"

/codon_start=1

/product="calcium-activated chloride channel protein 2"

/protein_id="AAD48398.1"

/db_xref="GI:5726288"

/translation="MGLRFGFVLLVLCILHQSNTSFIFKLNNGFEDIVVIDPSVPE
DEKLIQIEDMVTASTYLFETEKRFKNSILIPENKWPQYKPKHKKHAD
VIVAPTLPGRDEPTTKQTECKEYTHFTDLLGKKKNEYGPPGKLFHEWAHL
RWGFDYEDQFYRAKSKKIEATRCISAGISGRNRYKCGGSCLSRACRIDSTTKL
YKDCOFFPKVQTEKASIMFQSIDSVVFCNEKTHNQESLQIKCNFRSTWEVI
SNSEDFKNTIPMVTTPPPVPSLLKIRQIVCLVLDKSGSMGCKDLRNMNOAKHFL
LOTVENGSMVGMVHFDSTATIVNKLQIKSSDERNTLMAGLPTYPGLGTSICGIYA
FOVIGELHSDGSEVLLLDGDNATSCIDEVKSGGAIIVHFIALGRADAIVMS
KITGSHFYVSDAQNGLIDAFALTSQNTDLSQLESKGLTINSNMNDTVI
IDSTGKDTFFLTWNSLPSPISLWDPSTIMENFTVDATSKMAYLSIPGTAQVGTWA
YLIQAKNPETLITVTSAANSVPPITVNAKMKDVNSFPSPMIVYAEILQYVVP
LGANTATFESONGHTEVLELDNGAGADSKNDGVYSRVYATENGRIYSLAVRAHG
GANTARKLPPLNRAAYIPGVVNGEIANPPEIDETQTTLEDFTAGCGAFV
VSOVPSLPDQVPPSQITDLDTATHEKILFTWAPGNDFVKGVORYIIRISAIL
DLRSDFDALQVNTDLSREANSKESFAKPNISEENATHIFIAIKSIDKSNLTSK
VSNIAQVTLFIPOANPDIDPTPTPTTPDKSHNSGVNISTLVLSVIGSVIVNFILS
TTL"

BASE COUNT 1098 a 594 c 633 g 879 t

ORIGIN

alignment_scores::

Quality: 2872.50

Ratio: 3.775

Length: 913

Gaps: 8

Percent Similarity: 83.352

Percent Identity: 61.665

alignment_block:

US-09-049-696-41 x AF127035

Align seg 1/1 to: AF127035 from: 1 to: 3204

1 MetGlyProPheLysSerValPheIleLeuIleLeuHisLeuLeuG1 17
 29 ATGGGGTATTACAGAGGTTTGTCTTCTCTAGTCTGTGCTGCTGCA 78
 17 uGlyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34
 79 CCAGTCA...AATAGTCTCTTAAAGCTGAATATAATAGTCTTGAAG 125
 34 LyIleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
 126 ATATTGTCAATTGTTATAGATCCTAGTGTCCACAGAGATGAAAAATAAT 175
 51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
 176 GAACAAATAGAGGATATGGTGACTACAGCTTCTAGCTACTGTTGAAGC 225
 67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluT 84
 226 CACAGAAAAAGATTTTTTTCAAAAATGTATCTATATAATTCCGAGAGA 275
 84 hrTyrLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
 276 ATTGGAAGAAAACTTCAGTACAAAAGGCCAAAAACATCAAAACATAAA 325
 101 AsnAlaAspValLeuValAlaGluSerThrProProGlyAsnAspGluPr 117
 326 CATGCTGATGTTATAGTTTGACACCACTACACTCCAGGCTAGAGATGAAC 375
 117 oTyrThrGluGlnMetGlyAsnGlyCyluLysGlyGluArgIleHisL 134
 376 ATACACCAGCAGTTCACAGATGTGGAGAGAAAGCGGAATACATTCAC 425
 134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
 426 TCACCCCTGCCTTCTACTTGGAAAAACAAAAATGAATATGGACCACCA 475
 151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
 476 GGCAAACTGTTTGTCCATGAGTGGCTCACCTCCGGTGGGAGTGTTCGA 525
 167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsnGly...ArgIleG 183
 526 TGAGTACAAATCAAGATCAGCCTTCTACCGCTGAAGTCAAAAAAATCG 575
 183 InAlaValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLys 199
 576 AAGCAACAAGGTGTTCGCGAGGTATCTCTGGTGAAGAAATAGATTTATAAG 625
 200 CysGlnGlyGlySerCysTyrThrLysArgCysThrPheAsnLysValTh 216
 626 TGTCAAGGAGGACGTGCTTAGTAGAGCATGCGAATTTGATCTTCAAC 675
 216 rGlyLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrG 233
 676 AAAACTGTATGAAAAAGATGTCTCAATTTCTTCTCCTATAAAGTACAAACAG 725
 233 LuLysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPhe 249
 726 AAAAAGCATCATATGTTTATGCAAGATGATGATCTGTTCTGTTGAATTT 775
 250 CysThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLy 266
 776 TGTACAGGAAANACCATATACAGAGACTCCAGCCTTACAAACATAAA 825
 266 sCysAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheL 283
 826 GTGCAATTTTGAAGATACATGGGAGTCTATAGCAATTTCTGAGGATTTTA 875
 283 ysLysThrThrProMetThrThrGlnProProAsnProThrPheSerLeu 299
 876 AAAACACCATACCATGGTGACACACCATCTCCACATCTGCTTCATCATG 925

300 LeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLysSerGlyLeu 316
 316 rMetAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyClnLeuP 333
 333 heLeuLeuGlnThrValGlnLeuGlySerTrpValGlyMetValThrPhe 349
 349 1026 TCCTGCTGCACACTGTTGAAATGGATCCTGGGTGGGATGGTTCACTTT 1075
 1075 350 AspSerAlaAlaHisValGlnSerGluLeuIleGlnLeuAsnSerGlyLeu 366
 366 1076 GATAGTACTGCCACTATTGTAAATAAGCTAATCAAAATAAAAGCACTGA 1125
 1125 366 rAspArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerGlyGlyT 383
 383 1126 TGAAAGAACAACACTCATGGCAGGATTACCTPACATATACCTCTGGGAGAA 1175
 1175 383 hrSerIleCysSerGlyLeuArgSerAlaPheThrValIleArgLysLys 399
 399 1176 CTTCCATCTGCTCGGAATTAAATATGCATTTCAAGTGATTTGGAGACTA 1225
 1225 400 Tyr...ProThrAspGlySerGluIleValLeuLeuThrAspGlyGluAs 415
 415 1226 CATTCGCCAACTCGATGGATCCGAAGTACTGCTGCTGACTGATGGGAGGA 1275
 1275 415 pAsnThrIleSerGlyCysPheAsnGlnValLysGlnSerGlyAlaIleI 432
 432 1276 TAACACTCCAAAGTTCTTGATTTGATGAAGTGAAACAAGTGGGCCCATG 1325
 1325 432 IeHisThrValAlaLeuGlyProSerAlaAlaGlnLeuGluGluLeu 448
 448 1325 TFCATTTTATGCTTTGGGAAGAGCTGCTGATGAAGCAGTAATAGACATG 1375
 1375 448 SerLysMetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnValGlnAs 465
 465 1376 AGCAAGATAACAGGAGGAAGTCATTTTATGCTTTTCAGATGAAGCTCAGAA 1425
 1425 465 nAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAlav 482
 482 1426 CAATGGCCTCATATGATGCTTTTGGGGCTCTTTACATCATCAGAAATACTGATC 1475
 1475 482 alSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnAsn 498
 498 1476 TCTCCAGAAAGTCCCTTCAGCTCGAAGTAGAGGATTAACACTGAATAGT 1525
 1525 499 SerGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLysAs 515
 515 1526 AATGCTGGATGAACGACACTGTCTAATATTGATAGTACAGTGGGAAAGGA 1575
 1575 515 pThrLeuPheLeuIleThrTrpThrThrGlnProProGlnIleLeuLeuT 532
 532 1576 CAGGTTCCTTCTCATPACATGAACAGCTGCCCTCCAGCATTTTCTCTCT 1625
 1625 532 rAspProSerGlyGlnLysGlyGlyPheValValAspLysAsnThr 548
 548 1626 GGGATCCCGAGGAACATATGGAATAATTTCACAGTGGATGCACACTTC 1675
 1675 549 LysMetAlaTyrLeuGlnIleProGlyIleAlaLysValGlyThrTrpLys 565
 565 1676 AAAATGGCCTATCTCAGTATTTCCAGGAACATGCAAGGTGGGCACTTGGGC 1725
 1725 565 sTyrSerLeuGlnAlaSerSer.....GlnThrLeuThrLeuThrValT 580
 580 1726 ATACAATCTTCAAGCCAAAGCGCAACCCAGAAACATTAACCTATTACACTAA 1775
 1775 580 hrSerArgAlaSerAsnAlaThrLeuProProIleThrValThrSerLys 596
 596 1776 CTTCTCGAGCAGCAAAATTTCTTCTGTGCTCCAAATCAGTGAATGCTAAA 1825
 1825 597 ThrAsnLysAspThrSerLysPheProSerProLeuValValThrAlaAs 613

966 TGGAGCATGTCTTTCAGGAAGATCGCTCTCTTCCGAATGAATCAAGCAGCAG 1015
334 lnLeuPheLeuGlnThrValGlnLeuGlySerTrpValGlyMetVal 347
1016 AATTATCTTCATTCAAATATTGAAAGGGATCCCTTGGTTGGGATGGTT 1065
348 ThrPheAspSerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSe 364
1066 ACAATTGACAGTGTGCTGAATATCCGAATAATCTTAACAAAAAATACTGA 1115
364 rGlySerAspArgAspThrLeuAlaLysArgLeuProAlaAlaSerG 381
1116 TGATAAATCTTATGAAATATACCTGCAATCTGCCTCAAGAAGCTAATG 1165
381 lyGlyThrSerIleCysSerGlyLeuArgSerAlaPhe...ThrValIle 396
1166 GTGAACACTTCAATTTGTAGGGGGCTCAACAGCAGGATTTTCAGGCCAATTATT 1215
397 ArgLysLysTyrProThrAspGlySerGluIleValLeuLeuThrAspG 413
1216 CAGAGTCAACAGCAGCATTCTGGTTCTGAAATCATACTACTAACCGATGG 1265
413 yGluAspAsnThrIleSerGlyCysPheAsnGluValLysGlnSerGlyA 430
1266 GGAAGATAAATGAAATACACTCATGCTATTCAGGAGGTAAAAACAAGTGGTG 1315
430 laIleIleHisThrValAlaLeuGlyProSerAlaAlaGlnGluLeuGlu 446
1315 TGATCATTCACACCATTGCCCTGGGACCTTCGCTGCCAAAAGAACATGGAG 1365
447 GluLeuSerLysMetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnVa 463
1366 ACAGTGCAGATATGACAGGAGGACATCGTTTTATGCCAATAAAGACAT 1415
463 lGlnAsnAsnGlyIleLeuAspAlaPheGlyAlaLeuSerSerGlyAsnG 480
1416 A.....AATGCCCTTACTAATGCGCTTCAGTAGAATTTTCATCTAGAAGTG 1459
480 lyAlaValSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeu 496
1460 GCAGTATCAGTCAGCAGACTATTCAGTTGGAAGCAAGACCTTTGGCAATT 1509
497 GlnAsnSerGlnTrpMetAsnGlyThrValIleValAspSerThrValG 513
1510 ACAGAAAGAAATGGTAAATGTACGGTGCCTGTGATAGTACAAATGG 1559
513 yLysAspThrLeuPheLeuIleThrTrpThrGlnProGlnIleLe 530
1560 AAATGACACTTCTTTGTTGTTCACATGGACAATAAAAAAGCCAGAAATTC 1609
530 euLeuTrpAspProSerGlyGln.....LysGlnGlyGlyPheValVal 544
1610 TTCACAGAGTCCAAAGGAAGAAATAATAAACCTTCAGATTTCAAGAA 1659
545 AspLys...AsnThrLysMetAlaThrLeuGlnIleProGlyIleAlaL 560
1660 GATAGCTAAATATTCATTCCTCGCTCTCGNATACCTGGTATTGCAGA 1709
560 sValGlyThrTrpLysTyrSerLeu.....GlnAlaSerSerGlnT 574
1710 GACAGGTACTTGGACTTACAGCCTTCTAAATCAATCATGCGAGTCCCTCAA 1759
574 hrLeuThrLeuThrValThrSerArgAlaSerAsnAlaThrLeuProPro 590
1760 TACTAACAGTACAGCTGACCACTCCGACGAAGTCTCTACTACACCCCCA 1809
591 IleThrValThrSerLysThrAsnLysAspThrSerLysPheProSerPr 607
1810 GTAAGTCAACAGCTCAGTACATGAGTCAAAATACAGCAGTACCTAGCCCC 1859
607 oLeuValValTyrAlaAsnIleArgGlnGlyAlaSerProIleLeuArg 624

34 lyIleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
115 GTGTGGTCAATGGCCATTAAACCCAGTGTGCGGAGGAGCAAGGCTCATC 164
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
165 CCAAGCATAAAGGAATGTAACCTCAAGCCCTACCTACCTGTTGAAGC 214
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuLeuProGlu 84
215 CAGCCAAGGAAGAGTTATTTTCAGAAACATAAGCATATTAGTCCCGATGA 264
84 hrTyrLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
265 CTGGAGTCAAAACCTGAGTACTTATGCCCACAAACGAGATCGTATGAC 314
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
315 AAAGCAGACGTCATAGTTGCGGATCTCACCTGCAACATGGAGACGCC 364
117 oTyrThrGluGlnMetGlyAsnGlyGlyGluLysGlyGluArgIleHisL 134
365 CTACACCCCTTCAGTATGACAGTGTGGGACAGAGGACAGTACATACACT 414
134 euThrProAspPheIleAlaGlyLysLeuAlaGluTyrGlyProGln 150
415 TCACCTCCAACTCTCTACTACTGATAACTTTCGCTATCTACGGACCCCGA 464
151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
465 GGCAGAGTCTTTGTCATGAGTGGGCCCATCTCCGGTGGGAGTATTGGA 514
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSer...AsnGlyArgIleG 183
515 TCAGTATAACGTGGACCAACGCTTTCTACATGCTAGAAAGAACACTATAG 564
183 lnAlaValArgCysSerAlaGlyIleThrGlyThrAsnValLysLys 199
565 AAGCACAAGGTGTTCCACAGGATCACTGGCACCACCAATGTGTTCAACAC 614
200 CysGlnGlyGlySerCysTyrThrLysArgCysThrPheAsnLysValTh 216
615 TGTGAGAGAGCAACTGTGTCAAGGCGTCCGCGCTGACTCGAAGAC 664
216 rGlyLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrG 233
665 ACGGCTGTATGAACCAATGTACATTATCCAGACAAAATACAGACGG 714
233 luLysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPhe 249
715 CTGGGCGCTCCATATGTTTCATGCAAAACCTTAATCTGTGTTGAATTT 764
250 CysThrGluGlnAsnHisAsnGlySgluAlaProAsnLysGlnAsnGlnLy 266
765 TGCACAGAAAAAACCAATGCAAGGCGCCCAACCTACACAAACAAAAT 814
266 sCysAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheL 283
815 GTGCAATCGCAGAGCAGCTGGGATGTAATCAAGACGCTCTCTACTTTC 864
283 ysLysThrThrProMet.....ThrThrGlnProProAsnProThrPhe 297
865 AGATGCGCCCTCCATGAGAGACAGACAGAGCCCTCTCTCCACCTACATTT 914
298 SerLeuLeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLysSe 314
915 TCACTGCTCAAGTCCAGAGGCGAGTGTGTGTTGTTGGTGGATAATC 964
314 rGlySerMetAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyG 331
965 TGGAGCATGGCAAAAGAACCGCTCTATTTCGAATGAATCAAGCAGCAG 1014
331 lnLeuPheLeuLeuGlnThrValGluLeuGlySerTrpValGlyMetVal 347

1015 AACTGTACTTAACCAAAATGTGAAAAAGAGTCTATGTTGGATTAGTC 1064
348 ThrPheAspSerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSe 364
1065 ACATTGACACCGCTGCCACATCAAAATTTATCTAATAAAAAATAACGAG 1114
364 rGlySerAspArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerG 381
1115 TAGTAGTACTACCAAAAGATCCCGCAAACTCCCCCAACAGGCTTCTG 1164
381 lyGlyThrSerIleCysSerGlyLeuArgSerAlaPheThrValIleArg 397
1165 GTGGAATCTCAATTGGCATGACTCCAGGAGGATTTTCAGGCAATTAACC 1214
398 LysLys...TyrProThrAspGlySerGluIleValLeuLeuThrAspGl 413
1215 TCCAGTGACAGACACTCCCGTCTCAGATCGTATTGCTGCACAGATGG 1264
413 yGluAspAsnThrIleSerGlyCysPheAsnGluValLysGlnSerGlyA 430
1265 GGAAGATAATGGAATACGTTCTCTGTTGAGGCGCTCTCTCGCAGCGGTG 1314
430 laIleIleHisThrValAlaLeuGlyProSerAlaAlaGlnGluLeuGlu 446
1315 CCATCATCCACCATCGCTCTGGGCGCTTCGGTCCCGGAGAACTGGAG 1364
447 GluLeuSerLysMetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnVa 463
1365 ACTCTCGGACATGACAGAGGCGTTCGTTTCATGCAACAAAGACCT 1414
463 lGlnAsnAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnG 480
1415 A.....AACAGCCTTATCGATGCTTTTCAGTAGAATTTTCATCTACAAGTG 1458
480 lyAlaValSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeu 496
1459 CGAGGCTCTCCAGCAGGCTCTGCAGTTGGAGAGCAAGCCCTTCGATGTC 1508
497 GlnAsnSerGlnTrpMetAsnGlyThrValIleValAspSerThrValGl 513
1509 AGAGCAGGCGCATGGATAAAGGTCACAGTCTCTGGACAGTACCGTCGG 1558
513 yLysAspThrLeuPheLeuIleThrTrpThrThrGlnProGlnIleL 530
1559 CAACGACACGTTCTTGTATTACCTGGTAAAGGAGGCAAGAAATCA 1608
530 euLeuTrpAspProSerGlyGlnLys.....GlnGlyGlyPheValVal 544
1609 TTCCTCAAGATCCAAAGAGAAATAATACAACTCAGATTTCCAAGAT 1658
545 AspLys...AsnThrLysMetAlaTyrLeuGlnIleProGlyIleAlaLy 560
1659 GATAAATCAACATCCCGTCTGCTAGACTTCAAAATACCGGCGCAGTCA 1708
560 sValGlyThrTrpLysTyrSerLeuGlnAlaSer...SerGlnThrLeuT 576
1709 GACAGTACTTGGACTTACAGCATCCGGGTACCAAGTCTCAGTTGATTA 1758
576 hrLeuThrValThrSerArgAlaSerAsnAlaThrLeuProProIleThr 592
1759 CAATGACAGTGCACCTCGCAGCAAGAGTCCACCATGGAACCACTCTCTG 1808
593 ValThrSerLysThrAsnLysAspThrSerLysPheProSerProLeuVa 609
1809 GCTACTGCTCACATGAGTCAGAGCAGACGCCCGTACCTACCCGATCAT 1858
609 lValTyrAlaAsnIleArgGlnGlyAlaSerProIleLeuArgAlaServ 626
1859 TGTGTAGCAGCGGTGACCAAGGATTTTTCCTGCTGTTGGAGCCAATG 1908
626 alThrAlaLeuIleGluSerValAsnGlyLysThrValThrLeuGluLeu 642

```

1909 TCACAGCCCTCATAGAAGCTGACATGGACATCAAGTCACCTTTGGAGCTC 1959
643 LeuAspAsnGlyAlaGlyAlaAspAlaThrLysAspAspGlyValTyrSe 659
      |||||
1959 TGGACAATGGGCAGGTGCTGATACCTCTTAAAATGATGGCATCTACAC 2008
659 rArgTyrPheThrThrTyrAspThrAsnGlyArgTyrSerValLysVala 676
      |||||
2009 AGATACCTTTACAGATTATCATCGAAATGTTAGATACAGCTAAAGTGC 2058
676 rGalaLeuGlyGlyValAsnAlaAlaArgArgVal.....Ile 689
      |||||
2059 GT.....GTCCAGGCACAAAGAAACAAAACAGACGTGAGCTTA 2096
690 ProGlnGlnSerGlyAlaLeuTyrIleProGlyTyrPileGluAsnAspGl 706
      |||||
2097 AGACAGAACAACAAAGTCTTTATATATACCTGGCTATGTGCAAAATGGTAA 2146
706 uIleGlnTrpAspProArgProGluIleAsnLysAspAspValGlnH 723
      |||||
2147 AATTGTACTGAATCCACCAGACCATGTCCAAGAAGAACGCATAGAAG 2196
723 iSLysGlnValCysPheSerArgThrSerSerGlyGlySerPheValAla 739
      |||||
2197 CTACAGTGAAGACTTCAACAGAGTAACCTCTGGAGGGTCGTTACTGTG 2246
740 SerAspValProAsnAlaProIleProAsp.....LeuPh 751
      |||||
2247 TCTGGAGGCCCC.....CCTGATGGCACCAGCGCTCGTGTGT 2284
751 eProProGlyGlnIleThrAspLeuLysAlaGluIleHisGlyGlySerL 768
      |||||
2285 CCACCAAGTAAAGTCACAGACTGGAGGCTGAGTTATA.....GGTGATT 2331
768 euIleAsnLeuThrTrpThrAlaProGlyAspAspTyrAspHisGlyThr 784
      |||||
2332 ATATTACCTTACATGGAGCGCCCTCGCAAGGTTCTCGACAATGGAAAGA 2381
785 AlaHisLysTyrIleIleArgIleSerThrSerIleLeuAspLeuArgAs 801
      |||||
2382 GCACATAGATACATCATCAGATGAGCCAGCATCCTCTGATCTCCAAGA 2431
801 pLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuIleProLysG 818
      |||||
2432 AGATTTTAACATGCTACTTTTAAATGCTTCCAGTCTGATACCTAAAG 2481
818 luAlaAsnSerGluGluValPheLeuPheLysProGluAsnIleThrPhe 834
      |||||
2482 ANGCTGGCTCAAGAAGACATTAAATTCAAACCCAGAACTTTTAAATA 2531
835 GluAsnGlyThrAspPheIleAlaIleGlnAlaValAspLysValAs 851
      |||||
2532 GCAAAATGGCATCCAGCTCTACATTCCAATCCAGGCACACAATGAAGCCAG 2581
851 pLeuLysSerGluIleSerAsnIleAlaArgValSerLeuPheIleProP 868
      |||||
2582 TCTCACCTTCCTGAGGTCTCCAACATCGCA..... 2609
868 roGlnThrProProGluThrProSerProAspGluThrSerAlaProCys 884
      |||||
2610 ..CAGGCTGTCAAGCTTACTCTCTAGAAGATAGTATCTCTGCACCTGGT 2657
885 ProAsnIleHis....IleAsnSerThrIleProGlyIleHisIleLeu... 899
      |||||
2658 GATGATATTTCTGAATCTCTATGACAAATTTGGGGGTTAACTGTGATTT 2707
900 .....LysIleMetTrp 903
      |||||
2708 TAACCTCTATTTTAACTAGAGATAGATAGG 2738
seq_name: gb_ro:AF115852
seq_documentation_block:

```

114 GTGTGCTCATTTGCCATTAATACCCAGTGTGCCAGGACCAAGGCTCATC 163
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
164 CCAAGCATAAAGGAATGGTAACCAAGCTTCTACCTACCTGTTGAAGC 213
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuLeuProGluT 84
214 CAGCCAAGGAAGAGTTATTTTCAGGAACATAAGCATATAGTCCGATGA 263
84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
264 CCTGGAAGTCAAAATCTCAGTACTTAATGCCAAACAGAGATCGTACGAC 313
101 AsnAlaAspValLeuValAlaGluSerThrProProGlyAsnAspGluPr 117
314 AAAGCAGACGTCATAGTTGCGGATCCTCACCTGCAACATGGAGACGACC 363
117 oTyrThrGluGlnMetGlyAsnCysGlyGlyLysGlyGluArgIleHisL 134
364 CTACACCTTCAGTATGACAGCTGGGGAGCAGAGGACAGTACATACACT 413
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
414 TCACCTCAAACTTCTACTACTGATACTTCCGTATCTATGGACCCCGA 463
151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
464 GGCAGAGTCTTTGTCATGAGTGGGCCCATCTCCGGTGGGGAGTATTGA 513
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSer...AsnGlyArgIleG 183
514 TGAGTATAACGTGGACCGGCTTCTACATTTCTAGAAAGAACATATAG 563
183 InAlaValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLys 199
564 AAGCAACAAGGTGCTCCGCCAGCATCACAGCAAGAAGGTGTCACAG 613
200 CysGlnGlyGlySerCysTyrTrpLysArgCysThrPheAsnLysValTh 216
614 TGTACAGAGGACGCTGTGTGACAAAGGCGTCCCGCGCTGACTCGAAGAC 663
216 rGlyLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrG 233
664 ACGGCTGTATGAACCAATGTACATTTATCCAGACAAATACAGACAG 713
233 LuLysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPhe 249
714 CTGGGCGCTCCATATATGTTTCATGCAAAACCTCAATTCGTGGTTGAAT 763
250 CysThrGluGlnAsnHisAsnLysGluAlaProAsnLysGlnAsnGlnLy 266
764 TGCACAGAAAATAACCAATGCAGAAAGCCCAACACCTACAAAAAATA 813
266 sCysAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheL 283
814 GTGCATCGCAGACGACGTGGATGTAATCAAGCGCTCTCTGACTTTC 863
283 ysLysThrProMet...ThrThrGlnProProAsnProThrPhe 297
864 AGAATTCCTCCATGAGACGAACAGAGCGCTCTCCACCTACATTT 913
298 SerLeuLeuGlnIleGlyGlnArgIleValCysLeuValLeuAspLysSe 314
914 TCATTGCTCAAGTCCAGAAGCGAGTGTGTGCTGGTGGTGAATAATC 963
314 rGlySerMetAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyG 331
964 TGGAGCATGACAAAAGAACCGCTCTATTTCGAATGAATGAACGAGAG 1013
331 InLeuPheLeuGlnThrValGluLeuGlySerTrpValClyMetVal 347

1014 AACTGTACTTAACCTAAATTTGTGAAAAAGGAGTCTATGTTGGATTAGTC 1063
348 ThrPheAspSerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSe 364
1064 ACATTTGACAGCGCTGCCACATCCAAAATATATCTAATAAAAAAATACGAG 1113
364 rGlySerAspArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerG 381
1114 TAGTAGTGACTACCAAAAGATCACCGCAAACTCCCAACAGGCTACTG 1163
381 lyGlyThrSerIleCysSerGlyLeuArgSerAlaPheThrValIleArg 397
1164 GTGGAATTCATTTGCCATGAGCTCCAGCAGGATTTTCAGCAATTACC 1213
398 LysLys...TyrProThrAspGlySerGluIleValLeuLeuThrAspGl 413
1214 TCCAGTGACAGACGACTCCGGTCTGAGATCGTATTGCTGACAGATGG 1263
413 yGluAspAsnThrIleSerGlyCysPheAsnGluValLysGlnSerGlyA 430
1264 GGAAGATAATGAATAAGTTCTCTTTGAGGCGCTCTCTCGCAGCGGTG 1313
430 laIleIleHisThrValAlaLeuGlyProSerAlaAlaGlnGluLeuGlu 446
1314 CCATCATCCACACCATCGCTCTGGGGCTCTGCTGCCGAGAACTGGAG 1363
447 GluLeuSerLysMetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnVa 463
1364 ACTCTGTGACATGACAGAGGGCTTCTGTTCTATGCAACAAACATGT 1413
463 lclAsnAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnG 480
1414 A.....AGCAGCCTTATCGAGCGTTTCAGTAGGATTTCTACAAAGTG 1457
480 lyAlaValSerGlnArgSerIleGlnLeuGlnSerLysGlyLeuThrLeu 496
1458 CGAGCGTCTCCAGCAGCGCTCTGCAAGTTGGAGAGCAAGCCTTCAATGTC 1507
497 GlnAsnSerGlnTrpMetAsnGlyThrValIleValAspSerThrValGl 513
1508 AGAGCAGGGGATGGATAAAGTAGTACAGTACCTGTGGACAGTACGTCGG 1557
513 yLysAspThrLeuPheLeuIleThrTrpThrGlnProProGlnIleL 530
1553 CAACGACAGCTTCTTTGTTATCCTCGAGCGGTACAAAGCCAGCAATCA 1607
533 euLeuTrpAspProSerGlyGlnLys.....GlnGlyGlyPheValVal 544
1608 TCTTCAAGATCCAAAGAAAAAATAATATAACCTCAGATTTCCAGAT 1657
545 AspLys...AsnThrLysMetAlaTyrLeuGlnIleProGlyIleAlaLy 560
1659 GATGAACATAACATCCGGTCTCTCGACTTCAATATACGGGCACTGCAGA 1707
560 sValGlyThrTrpLysTyrSerLeuGlnAlaSer...SerGlnThrLeuT 576
1702 GACAGGTACTTGGACTTACAGCATCACGGGTACCAAGTCTCAGTTGATTA 1757
576 hrLeuThrValThrSerArgAlaSerAsnAlaThrLeuProProIleThr 592
1758 CAATGACAGATGACCACTCGACCAAGTCCCAACCATGGAACCACTCTCG 1807
593 ValThrSerLysThrAsnLysAspThrSerLysPheProSerProLeuVa 609
1808 GCTACCGCTCAGATGAGTCAGAGCAGACGACAGTACCTAGCCGATGAT 1857
609 lValThrAlaAsnIleArgGlnGlyAlaSerProIleLeuArgAlaServ 626
1858 TGTGTATGACGGGTACCAAGGGTTTTTGCCTGTTCTAGGAGCAATG 1907
626 alThrAlaLeuIleGlnSerValAsnGlyLysThrValThrLeuGluLeu 642
1908 TCACAGCCATCATAGAACTGACATGACATCAAGTCAACCTCGGAGCTA 1957

Align seg 1/1 to: BC008147 from: 1 to: 3137

[illegible][illegible]

17 uclyAlaLeuSerAsnSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34
97 GAACACA...GAGAGCTCCATGGTGCATCTCAACAGCAAGATACGAGG 143
34 lyIleValValAlaIleAspProAsnValProGluAspGluThrLeuIle 50
144 GTGTGTCATTGCCATTAAACCCAGTGTGCCGAGGACGAAGGCTCATC 193
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGluAl 67
194 CCAAGCAATAAGGAATGGTAACCTCAAGCCCTACCTACCTGTTTGAAGC 243
67 aThrGlyLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluT 84
244 CAGCAAGGAAGAGTTATTTTCAGAAACATAAGCATATTAGTCCCGATGA 293
84 hrTrpLysThrLysAlaAspTyrValArgProLysLeuGluThrTyrLys 100
294 CCTGGAAGTCAAAACCTCAGTACTTAATGCCAAACAGAGATCGTATGAC 343
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
344 AAAGCAGACGTCATAGTTCGCGATCCTCACCTGCAACATGGAGAGCACC 393
117 oTyrThrGluGlnMetGlyAsnGlyGlyGluLysGlyGluArgIleHisL 134
394 CTACACCTTCAGTAGGACAGTGTGGGACAGAGGACATACATACACT 443
134 euThrProAspPheIleAlaGlyLysLeuAlaGluTyrGlyProGln 150
444 TCACCTCCAAACTTCCTACTCATGATAACTTCGGTATCTACGGACCCCGA 493
151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
494 GGCAGAGTCTTTGTCCATGAGTGGGCCCATCTCCGGTGGGAGTATTGA 543
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSer...AsnGlyArgIleG 183
544 TGAGTATTAACGTGGACAGCCCTTCTACATGCTAGAAAGAACACTATAG 593
183 InAlaValArgCysSerAlaGlyIleThrGlyThrAsnValValLysLys 199
594 AAGCAACAAGGTGTCCACCAGGATCACTGGCACCACCAATGTGTTTCAAC 643
200 CysGlnGlyGlySerCysTyrThrLysArgCysThrPheAsnLysValTh 216
644 TGTGAGAGAGCAACTGTGTCAACAGGCGCTCCGCGGTGACTCGAAGAC 693
216 rGlyLeuTyrGluLysGlyCysGluPheValLeuGlnSerArgGlnThrG 233
694 ACGGCTGTATGACCAACCAATGTACATTTATCCAGACAAATAACAGACG 743
233 luLysAlaSerIleMetPheAlaGlnHisValAspSerIleValGluPhe 249
744 CTGGGCGCTCCATATATGTTATGCAAAACCTTAATTCGTGTTGAATT 793
250 CysThrGluGlnHisAsnLysGluAlaProAsnLysGlnAsnGlnLy 266
794 TGCACAGAAAAAACCAATGCGAAGCCCAACACCTACAAAAACAAAT 843
266 sCysAsnLeuArgSerThrTrpGluValIleArgAspSerGluAspPheL 283
844 GTGCAATCCGACAGACGCTGGGTATTAATCAAGACGCTCTGCTGACTTTC 893
283 ysLysThrThrProMet.....ThrThrGlnProProAsnProThrPhe 297
894 AGAATGCCCTCCCATGATGAGACAGAACAGACCCCTCCTCCACCTACATT 943
298 SerLeuLeuGlnIleGlyValArgIleValCysLeuValLeuAspLysSe 314
944 TCACCTGCTCAAGTCCAGAGCGGAGTGTGCTTGGTGTCTGGATAAATC 993
314 rGlySerMetAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyG 331

1043 TGGAGCATGGCAAAAGAGCCGCTTATTTCGAATGAATCAAGCAGCAG 1043
331 InLeuPheLeuLeuGlnThrValGluLeuGlySerTrpValGlyMetVal 347
1044 AACTGTACTTAACTCAAAATTTGGAAAGGAGCTCTATGTTGGATTAGTC 1093
348 ThrPheAspSerAlaAlaHisValGlnSerGluLeuIleGlnIleAsnSe 364
1094 ACATTTGACAGCGCTGCCACATCCAAATATCTAATAAAAAAATACGAG 1143
364 rGlySerAspArgAspThrLeuAlaLysArgLeuProAlaAlaAlaSerG 381
1144 TAGTAGTACTACCAAAAGATCACCGCAAAACCTCCCAACAGGCTTCTG 1193
381 lyGlyThrSerIleCysSerGlyLeuArgSerAlaPheThrValIleArg 397
1194 GTGGAATTCATTTGCCATGGACTCCAGGCAGGATTTTCAGCAATTACC 1243
398 LysLys...TyrProThrAspGlySerGluIleValLeuLeuThrAspGl 413
1244 TCCAGTGACACAGACACTTCGGGTTCTGAGATCGTATTGCTGACAGATGG 1293
413 yGluAspAsnThrIleSerGlyCysPheAsnGluValLysGlnSerGlyA 430
1294 GGAGATAATGGAATACGTTCTGCTTTGAGCCGCTCTCTCCGACGGTG 1343
430 IalIleHisThrValAlaLeuGlyProSerAlaAlaGlnGluLeuGlu 446
1344 CCATCATCCACACCATCGCTCTGGGCGCTTCGGCTGCCGAGAACTGGAG 1393
447 GluLeuSerLysMetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnVa 463
1394 ACTCTGTCGGACATGACAGAGGGCTTCGTTTCATGCCAACAAAGACCT 1443
463 lGlnAsnAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnG 480
1444 A.....AACAGCCTTATCGATGCTTTTCAGTAGAATTTCACTACAAGTG 1487
480 lyAlaValSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeu 496
1488 GCAGCGTCTCCAGCAGGCTCTGCAAGTGGAGAGCAAAAGCCTTCGATGTC 1537
497 GlnAsnSerGlnTrpMetAsnGlyThrValIleValAspSerSerValGl 513
1538 AGACAGGGGCGATGGATAAACGGTACAGTACCTCTGGACAGTACCGTCGG 1587
513 yLysAspThrLeuPheLeuIleThrTrpThrGlnProGlnIleL 530
1588 CAACGACACGTTCTTTGTTTACCTCGATGGTGTAAAAAGCCAGAAATCA 1637
530 euLeuTrpAspProSerGlyGlnLys.....GlnGlyGlyPheValVal 544
1637 TCTTCAAGATCCAAAGAGAAAAAATATACAACCTCAGATTTCACAGAT 1687
545 AspLys...AsnThrLysMetAlaTyrLeuGlnIleProGlyIleAlaLy 560
1688 GATAACATAACATCCGGTCTCTAGACTTCAATACCGGGCAGCTGCAGA 1737
560 sValGlyThrTrpLysTyrSerLeuGlnAlaSer...SerGlnThrLeuT 576
1736 GACAGTACTTGGACTTACAGCATCACGGGTACCAGTCTCAGTTGATTA 1787
576 hrLeuThrValThrSerArgAlaSerAsnAlaThrLeuProProIleThr 592
1788 CAATGACAGTGACCACTCTGAGCAAGAAGTCCCAACCATGGAACCACTCCTG 1837
593 ValThrSerLysThrAsnLysAspThrSerLysPheProSerProLeuVa 609
1838 GCTACTGCTCACATGATGACAGCAGCAGCCCACTACCTACCGGATGAT 1887
609 lValThrAlaAsnIleArgGlnGlyAlaSerProIleLeuArgAlaSerV 626

```
1888 TGTGTACGACGGGTACGCCAGGATTTTTCCTGCTTCTGGAGCCAATG 1937
626 alThrAlaLeuLeuSerValAsnGlyLysThrValThrLeuGluLeu 642
|||||
1938 TCACAGCCTCATAGAACTGAACATGACATCAAGTCACCTGGAGCTC 1987
643 LeuAspAsnGlyAlaGlyAlaAspAlaThrLysAspAspGlyValTyrSe 659
|||||
1988 TGGGCAATGGGCGAGGTGCTGATCTTTAAATGATGTCATCTACAC 2037
659 rArgTyrPheThrThrTyrAspThrAsnGlyArgTyrSerValLysVala 676
|||||
2038 AAGATACCTTACAGATTATATGGAATAGTAGATACAGCTAAAGTGC 2087
676 rGAlaLeuGlyGlyValAsnAlaAlaArgArgVal.....Ile 689
||
2088 GT.....GTCCAGGCAGAAACAAACAAACACGACGCTGAGCTTA 2125
690 ProGlnGlnSerGlyAlaLeuTyrIleProGlyTrpIleGluAsnAspG1 706
|||||
2126 AGACAGAAACAAGCTCTTTATATATACCTGGCTATGTGAAATGTGTA 2175
706 uileGlnTrpAsnProArgProGluIleAsnLysAspAspValGlnH 723
|||||
2176 AATTGTACTGAATCCACCCAGACAGATGTCACAGAAAGACCCATAGAAG 2225
723 isLysGlnValCysPheSerArgThrSerSerGlyGlySerPheVala 739
|||||
2226 CTACAGTGGAGAGCTTCAACAGAGTAACTCTGGAGGCTGTTTACTGTG 2275
740 SerAspValProAsnAlaProIleProAsp.....LeuPh 751
|||||
2276 TCTGGAGCGGCC.....CTGATGGCGACCACTCGTGTGT 2313
751 eProProGlyGlnIleThrAspLeuLysAlaGluIleHisGlyGlySerL 768
|||||
2314 CCCACCAAGTAAAGTCACAGACCTGGAGGCTGAGTTATA...GGTGATT 2360
768 euileAsnLeuThrTrpThrAlaProGlyAspAspThrAspHisGlyThr 784
|||||
2361 ATATTACCTTACATGACGCGCCCTGGCAAGGTTCTCGACAATGGAAGA 2410
785 AlaHisLysTyrIleIleArgIleSerThrSerIleLeuAspLeuArgAs 801
|||||
2411 GCACATAGATACATCATCAAGATGAGCCAGCATCTCTGATCTCCAAGA 2460
801 pLysPheAsnGluSerLeuGlnValAsnThrThrAlaLeuIleProLysG 818
|||||
2461 AGATTTTACAATGCTACTTTAGTGATGCTTCCAGTCTGATACCTAAAG 2510
818 luAlaAsnSerGluGluValPheLeuPheLysProGluAsnIleThrPhe 834
|||||
2511 AAGCTGGCTCAAAAGAAACATTTAAATTCAAACCAACAACTTTTAAATA 2560
835 GluAsnGlyThrAspLeuPheIleAlaIleGlnAlaValAspLysValas 851
|||||
2561 GCNAATGGCATTCAGCTTTACATGTCATCAATCCAGGAGCAATGAGCCAG 2610
851 pLeuLysSerGluIleSerAsnIleAlaArgValSerLeuPheIleProp 868
|||||
2611 TCTCACTCTGAGGTCTCAACATCGCA..... 2638
868 roGlnThrProGluThrProSerProAspGluThrSerAlaProCys 884
|||||
2639 ..CAGGCTCTCAAGCTTACTTCTCTAGAGATAGATATCTCTGCACTGGGT 2686
885 ProAsnIleHisIleAsnSerThrIleProGlyIleHisIleLeu..... 899
|||||
2687 GATATTTCTGCAATCTGTATGACATTTGGGGTTAACTGTGATTTTAA 2736
900 .....LysIleMetTrp 903
|||||
2737 CTCATTTTAAACTAGAGATAGATAGG 2764
```

```
seq_name: gb_ro:AF108501
seq_documentation_block: 3058 bp mRNA ROD 23-NOV-1999
LOCUS AF108501 Ca(2+)-sensitive chloride channel 2 (Cacc) mRNA,
DEFINITION complete cds.
ACCESSION AF108501
VERSION AF108501.1 GI:6465944
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 3058)
AUTHORS Lee,D., Ha,S., Kho,Y., Kim,J., Cho,K., Baik,M. and Choi,Y.
TITLE Induction of mouse Ca(2+)-sensitive chloride channel 2 gene during
involution of mammary gland
JOURNAL Biochem. Biophys. Res. Commun. 264 (3), 933-937 (1999)
MEDLINE 20012773
PUBMED 10544033
REFERENCE 2 (bases 1 to 3058)
AUTHORS Lee,D., Ha,S., Kho,Y. and Choi,Y.
TITLE Direct Submission
JOURNAL Submitted (22-NOV-1998) Department of Animal Science & Technology,
Seoul National University, Seodun 103, Suwon, Kyungki 441-744,
Republic of Korea
FEATURES
Source Location/Qualifiers
1..3058
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_type="epithelial"
/tissue_type="mammary"
1..3058
/gene="Cacc"
17..2725
/gene="Cacc"
/feature="CCLCA2; regulated by lactogenic hormones; increased
expression during apoptosis of mammary epithelial cells"
/codon_start=1
/product="Ca(2+)-sensitive chloride channel 2"
/protein_id="AAFI2731.1"
/db_xref="GI:6465945"
/translational="MVPGLQVLLFLTLHLQNTSSVMVHLNSNGYGVIAINPSVPE
DERLIPSIKEMVTQASTYLEASGRVYFRNISIILVPMTWKSKSEYILMPKRESYDAD
VIVADPLHQHDDPYTLQYQCGDRGYIHFTNPLTLTDLNRIYGPGRVVEHWAHL
RWGVEDYNNVDRPFYISRNTIEATCSASITGKVVHECQRCGCVTRACRDSKTP
YEPCTFIPDKIOTAGASIMFMONLNSVVECTENNHNENAPNLONKMCNRRSTWDVI
KASADFNQSPMRGTEAPPPTFSLKSRVYVCLVLDKSGMDKEDRLRMQAAEL
YLTQIVEKSWGLVTFDSAAHQNLIKITSSDYOKITANLPQATGTSICHGLQ
AGQATITSDQSTSGSEIVILLTGDGNGISSCFEAVSRSGAIIHTALGPSAAELET
LSDMTGLGRFYANKHVSLLIDAFSRISSTSGVSOQAOLIESKAFNVACAGINSTVL
VDSVTGNDTFVITVTQKPEIILQDPKGGYITSDODDELNIRARLIQIPCTAETG
TWTSITGKTSQITMTVTTRASPTMEPLATAHMSQSTAOYPSRMIVYVRSQFEL
PVLGADVATIAEHGHOVTELEWDNAGADTVKNDGIVTRYFTDTHNGRYSIKVRV
QARKNKARLSLRKNSLPIGVYKNGKIVNPNRPDQVEEALEATEVDFNRTVSGS
LTVSGAPPDGDHARVPPKSTVDLEAEFFIGDYIQLTWTAPKVDLDRGHRYIIRVSG
HPLGLQEDFNNTLVNASSLIPKEAGSKETFKPETFKTIANDTQLYIAIQAYNEAGL
TSEVSNIAQVAKFTSLSDSISALGADISATISMTVWGLAVIFNSIL"
```

```
BASE COUNT 951 a 706 c 670 g 731 t
ORIGIN
alignment_scores:
Quality: 2353.50 Length: 913
Ratio: 3.324 Gaps: 14
Percent Similarity: 77.547 Percent Identity: 52.464
alignment_block:
US-09-049-696-41 x AF108501
Align seg 1/1 to: AF108501 from: 1 to: 3058
```

[illegible]

914 TCATTGCTCAAGTCCAGAAGCGAGTGGTGTGCTTGGTGTGATAAATC 963
 314 rGlySerMetAlaThrGlyAsnArgLeuAsnArgLeuAsnGlnAlaGlyG 331
 964 TGGAGACATGGACAAAGAAGACCGCTCTATTTCGAATGAATCAAGACGAC 1013
 331 InLeuPheLeuGlnThrValGluLeuGlySerTrpValGlyMetVal 347
 1014 AACTGTACTTAATCAAAATTTGGAAAAGGAGTCTATGGTTGGATTAGTC 1063
 348 ThrPheAspSerAlaAlaHisValGlnSerGluLeuLeuGlnLeuAsnSe 364
 1064 ACATTGTGACGGGTGCCACATCCAAAATTTATCTAATAAAAAATACGAG 1113
 364 rGlySerAspArgAspThrLeuAlaLysArgLeuProAlaAlaSerG 381
 1114 TAGTAGTGACTACCAAAAGATACCGCAAACTCCCCACACAGCGTACTG 1163
 381 yGlyThrSerIleCysSerGlyLeuArgSerAlaPheThrValIleArg 397
 1164 GTGGAACTTCAATTGGCATGCGACTCCGAGCAGGATTTCAGGCAATTACC 1213
 398 LysLys...TyrProThrAspGlySerGluIleValLeuLeuThrAspG 413
 1214 TCCAGTGACACGACACTTCGGTCTTCGATCGTATTGCTGCACATGG 1263
 413 yGluAspAsnThrIleSerGlyCysPheAsnGluValLysGlnSerGlyA 430
 1264 GGAAGATAATGGAAVAGTCTCTGTTGAGCGCGTCTTCGAGCGGGT 1313
 430 IaIleHisThrValAlaLeuGlyProSerAlaAlaGlnGluLeuGlu 446
 1314 CCATCATCCACACCATCGCTCTGGGGCTTCGTGTCGCCGAGAACTGGAG 1363
 447 GluLeuSerLysMetThrGlyGlyLeuGlnThrTyrAlaSerAspGlnVa 463
 1364 ACTCTGTCGACATCATCAGAGGAGCGTTCTGTTCTATGCCAACAAACATGT 1413
 463 IglInAsnAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnG 480
 1414 A.....AGCAGCCTATCGACGCTTCAGTAGGATTTCATACAAAGT 1457
 480 yAlaValSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeu 496
 1458 GCAGCGTCTCCAGCAGGCTCTGCAGTTGGAGACAAAGCCTTCAATGTC 1507
 497 GlnAsnSerGlnTrpMetAsnGlyThrValIleValAspSerThrValG 513
 1508 AGACAGAGGGGATGGATAAACAGTACAGTACTTGTGGACAGTACCGTCGG 1557
 513 yLysAspThrLeuPheLeuLeuThrTrpThrThrGlnProGlnIle 530
 1558 CAACGACAGCTCTTTGTTATCCTCGGCGGTACAAAAGCCAGAAATCA 1607
 530 euLeuTrpAspProSerGlyGlnLys.....GlnGlyGlyPheValVal 544
 1608 TTCITCAAGATCCAAAGGAAAAAATATATACTTCAGATTTCCAAGAT 1657
 545 AspLys...AsnThrLysMetAlaTyLeuGlnIleProGlyIleAlaTy 560
 1658 GATGAACATAACATCCCGTCTGCTCGACTTCAAATACCGGGCACTGCAGA 1707
 560 sValGlyThrTrpLysTySerLeuGlnAlaSer...SerGlnThrLeuTr 576
 1708 CACAGGTACTTTGGACTTACAGCATCACGGGTACCAAGTCTCAGTTGATTA 1757
 576 hrLeuThrValThrSerArgAlaSerAsnAlaThrLeuProPheThr 592
 1758 CAATGACAGTGACCACCTCGGACAGAGAGTCCCCACATGGGAACCACTCC 1807
 593 ValThrSerLysThrAsnLysAspThrSerLysPheProSerProLeuVa 609

1808 GCTACCGCTCACATGAGTCAGCAGCACGCCAGTACCTAGCCGGATGAT 1857
609 lvaltyrAlaAsnIleArgGlnGlyAlaSerProIleLeuArgAlaSerV 626
1858 TGTGTATGTACGGGTACGCCAAGGGTTTTTGCCTGTCTAGGAGCCGATG 1907
626 alThrAlaLeuIleGluSerValAsnGlyLysThrValThrLeuGluLeu 642
1908 TCACAGCCATCATAGAGCTGACATGACATCAAGTCACCCCTGGAGCTA 1957
643 LeuAspAsnGlyAlaGlyAlaAspAlaThrLysAspGlyValTyrse 659
1958 TGGACAATGGGCAGGTGCTGATACCTGTTAAAAATGATGTCATACAC 2007
659 rArgTyrPheThrThrTyrAspThrAsnGlyArgTyrSerValLysVala 676
2008 AGATACCTTTACAGATTATCATGGAATGTTAGATACAGCTAAAAGTGC 2057
676 rgAlaLeuGlyGlyValAsnAlaAlaArgArgValIleProGlnGln 692
2058 GTGTCCAGGCAAGAAAAACAAGCCAGACTAAGC...TTAAGACAGAAG 2104
693 SerGlyAlaLeuTyrIleProGlyTyrIleGluAsnAspGluIleGlnTr 709
2105 AACAAAGTCTTTATATATACCTGCTATGTGGAATGTTAAATTTGTACT 2154
709 pAsnProProGluProGluIleAsnLysAspValGlnHisLysGlnV 726
2155 GAATCCACCAGACCAAGATGTCACAGAAGAAGCCATAGAAGCTACAGTGG 2204
726 alCysPheSerArgThrSerSerGlySerPheValAlaLaserAspVal 742
2205 AGGACTTCAACACAGTAACTCTGGAGGTCATTGACTGTGCTGGAGCG 2254
743 ProAsnAlaProIleProAsp.....LeupheProProgl 754
2255 CCC.....CCTGATGGCGACCATGCTCGTGTGTTCCCAACAAG 2292
754 yGlnIleThrAspLeuLysAlaGluIleHisGlyGlySerLeuIleAsnL 771
2293 TAAAGTCACAGACCTGGAGGCTGAGTTTATA...GGTGATTACATTCAAC 2339
771 euThrTrpThrAlaProGlyAspTyrAspHisGlyThrAlaHisLys 787
2340 TTACATGGAGCGCCCTGGCAAGGTTCTCGATAAGGGAAGAGCGCATAGA 2389
788 TyrIleIleArgIleSerThrSerIleLeuAspLeuArgAspLysPheAs 804
2390 TACATCATCAGAGTGAGCCAGCATCTCTGGGTCTCCAGAAGATTTTAA 2439
804 nGluSerLeuGlnValAsnThrThrAlaLeuIleProLysGluAlaAsnS 821
2440 CAATGTTACTTTAGTGAATGCTTCCAGTCTGATACCTAAAGAGGCTGGCT 2489
821 exGluGluValPheLeuPheLysProGluAsnIleThrPheGluAsnGly 837
2490 CAAAGAAACATTAAATTCAAACCAGAACTTTTAAATAGCAATGAC 2539
838 ThrAspLeuPheIleAlaIleGlnAlaValAspLysValAspLeuLysSe 854
2540 ACCCAGCTTACATTGCAATCCAGGCATACATGAAGCTGGTCTCACCTC 2589
854 rGluIleSerAsnIleAlaArgValSerLeuPheIleProProGlnInThr 871
2590 TGAGGTCTCCACATTGCACAGGCTGTCAAGTTT..... 2623
871 roProGluThrProSerProAspGluThrSerAlaProCysProAsnIle 887
2624ACTTCTCTAAGATAGTATCTCTGCACCTGGGTGCTGATATT 2665
888 His.....IleAsnSerThrIleProGlyIleHisIleLeu 899
2666 TCTGCAATCTCTATGACAGTTTGGGGGTAGCTGTGATT 2704

Align seg 1/1 to: AAG73854 from: 1 to: 552

1 GACACAGCAAAATCCCGCCCTCTGGTATGCAAAATATTCGCCA 50
|||||
238 AspThrSerLysPheProSerProLeuValValTyrAlaAsnIleArgG1 254
51 AGGAGCCTCCCAATTCCTCAGGCGCCAGTGTACAGCCCTGATTGAATCAG 100
|||||
254 nGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerV 271
101 TGAATGGAAGAACAGTACCTTGGAACTACTGTGATAATGACAGCGTCT 150
|||||
271 alAsnGlyLysThrValThrLeuGluLeuAspAsnGlyAlaGlyAla 287
151 GATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTTCACTATATCA 200
|||||
288 AspAlaThrLysAspAspGlyValTyrSerArgTyrPheThrThrTyrAs 304
201 CACCAATGCTAGATACAGTGTAAAGTGGCGGCT 234
|||||
304 pThrAsnGlyArgTyrSerValLysValArgAla 315

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAG75614

seq_documentation_block:
ID AAG75614 standard; Protein; 869 AA.

XX AC AAG75614;
XX XX
DT 03-SEP-2001 (first entry)
XX Human colon cancer antigen protein SEQ ID NO:6378.
XX Human: colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX Homo sapiens.
XX WO200122920-A2.
XX 05-APR-2001.
XX 28-SEP-2000; 2000WO-US26524.
XX 29-SEP-1999; 99US-0157137.
XX 03-NOV-1999; 99US-0163280.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPT; 2001-235357/24.
XX N-PSDB; AAH35019.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 11; Page 7851-7854; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis

CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 869 AA;

alignment_scores:
Quality: 392.00 Length: 78
Ratio: 5.026 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-12 x AAG75614 ..

Align seg 1/1 to: AAG75614 from: 1 to: 869

1 GACACAGCAAAATCCCGCCCTCTGGTATGCAAAATATTCGCCA 50
|||||
555 AspThrSerLysPheProSerProLeuValValTyrAlaAsnIleArgG1 571
51 AGGAGCCTCCCAATTCCTCAGGCGCCAGTGTACAGCCCTGATTGAATCAG 100
|||||
571 nGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerV 588
101 TGAATGGAAGAACAGTACCTTGGAACTACTGTGATAATGACAGCGTCT 150
|||||
588 alAsnGlyLysThrValThrLeuGluLeuAspAsnGlyAlaGlyAla 604
151 GATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTTCACTATATCA 200
|||||
605 AspAlaThrLysAspAspGlyValTyrSerArgTyrPheThrThrTyrAs 621
201 CACCAATGCTAGATACAGTGTAAAGTGGCGGCT 234
|||||
621 pThrAsnGlyArgTyrSerValLysValArgAla 632

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74824

seq_documentation_block:
ID AAB74824 standard; Protein; 914 AA.

XX AC AAB74824;
XX XX
DT 13-JUN-2001 (first entry)
XX Human ICACC-1 protein sequence.

XX ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
XX interleukin 9 induced calcium activated chloride channel; IL-9;
XX calcium activated chloride channel; anti-allergic; anti-asthmatic;
XX anti-inflammatory; immunomodulatory; cystic fibrosis;
XX inflammatory bowel disease; autoimmune disease.

XX Homo sapiens.

XX WO9944620-A1.

XX 10-SEP-1999.

XX 03-MAR-1999; 99WO-US04703.

XX 03-MAR-1998; 98US-0076815.

XX (MAGA-) MAGAININ PHARM INC.

XX Holroyd KJ, Levitt RC, Maloy WL, Loughed J, McLane M;

XX Nicolaides NC, Zhou Y, Dong Q;

XX WFI; 1999-550979/46.

XX DR

DR N-PSDB; AAF81927.
XX
PT New nucleic acid encoding calcium activated chloride channel, used to
XX identify, e.g. specific modulators for treating atopic allergy -
XX
XX Claim 11; Fig 4B; 75pp; English.
XX
XX The present sequence represents the human interleukin 9 (IL-9) induced
XX calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
XX have anti-allergic, anti-asthmatic, anti-inflammatory and
XX immunomodulatory activities. Compounds (A) that downregulate ICACC are
XX used to alleviate asthma (or more generally atopic allergy), while those
XX (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
XX inflammatory bowel disease (IBD) (or other autoimmune diseases).
XX Measuring levels of ICACC-1 can be used in the diagnosis of asthma
XX (increased levels) or IBD (reduced levels), also for monitoring
XX treatment of these conditions. The ICACC proteins can be used:
XX (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
XX reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
XX (A) to disrupt bonding between ICACC and its ligands; and (iii) to
XX identify modulators and binding partners. ICACC polynucleotides can be
XX used to generate transgenic animals or recombinant cells, used to screen
XX for antagonists, also as a source of therapeutic antisense agents or
XX diagnostic probes (for quantifying mRNA expression, e.g. for
XX identification of modulators).

SQ Sequence 914 AA;

alignment_scores:
Quality: 392.00 Length: 78
Ratio: 5.026 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-12 x AAB74824 ..

Align seg 1/1 to: AAB74824 from: 1 to: 914

1 GACACCAGCAAAATCCCGCCCTCTGCTAGTTATGCAAAATATTCGCCA 50
|||||
600 AspThrSerLysPheProSerProLeuValValTyrAlaAsnIleArgG1 616
51 AGGACCTCCCAATCTCAGGGCCAGTGTACAGCCCTGATGATCAG 100
|||||
616 nGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerV 633
101 TGAATGGAAAAACAGTTACCTTGGAACTACTGGATAATGGAGCGTGTCT 150
|||||
633 AlaAsnGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAla 649
151 GATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCAACAATATGA 200
|||||
650 AspAlaThrLysAspAspGlyValTyrSerArgTyrPheThrThrTyrAs 666
201 CACGAATGCTAGATACAGTGTAAAGTCGGGGCT 234
|||||
666 pThrAsnGlyArgTyrSerValLysValArgAla 677

seq_name: /SIDS2/gcgdata/geneseqp/AA2001.DAT:AA24514

seq_documentation_block:

ID_AA24514 standard; Protein; 914 AA.

XX AA24514;

XX 12-OCT-2001 (first entry)

XX C902P predicted amino acid sequence.

XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW immunogenic; gene therapy; vaccine; colon cancer.

XX

OS Homo sapiens.
XX
PN WO200149716-A2.
XX
PD 12-JUL-2001.
XX
XX 29-DEC-2000; 2000WO-US35596.
XX
XX 30-DEC-1999; 99US-0476296.
PR 10-JAN-2000; 2000US-0480321.
PR 15-FEB-2000; 2000US-0504629.
PR 06-MAR-2000; 2000US-0519444.
PR 19-MAY-2000; 2000US-0575251.
PR 29-JUN-2000; 2000US-0609448.
PR 28-AUG-2000; 2000US-0649811.
XX

(CORI-) CORIXA CORP.

XX Xu J. Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jlang Y;
XX

DR WPI; 2001-441847/47.

XX Colon tumor associated proteins and nucleic acids useful for the
XX prevention, diagnosis and treatment of colonic cancer -

Claim 2; Page 440-443; 472pp; English.

XX The present invention describes colon tumour associated proteins (I) and
XX the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX (II) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate colon tumour associated protein (TCAP)
XX expression, such as colonic cancer. For example, (I) and (II) may be
XX used to treat disorders associated with decreased expression by
XX rectifying mutations or deletions in a patient's genome that affect the
XX activity of TCAPs by expressing inactive proteins or to supplement the
XX patients own production of them. Additionally, (II) may be used to
XX produce the TCAP proteins, by inserting the nucleic acids into a host
XX cell culturing the cell to express the protein. (II) and its
XX complementary sequences may also be used as DNA probes in diagnostic
XX polymerase chain reaction (PCR) and hybridisation assays to detect and
XX quantitate the presence of similar nucleic acids in samples, and
XX therefore which patients may be in need of restorative therapy. (I) may
XX also be used as antigens in the production of antibodies against TCAPs
XX and in assays to identify modulators of TCAP expression and activity.
XX Anti-(I) antibodies and antagonists may also be used to down regulate
XX TCAP expression and activity. The anti-(I) antibodies may also be used
XX as diagnostic agents for detecting the presence of TCAPs in samples
XX (e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512
XX and AA24494 to AA24523 represent nucleotide and amino acid sequences
XX given in the exemplification of the present invention.

SQ Sequence 914 AA;

alignment_scores:

Quality: 392.00 Length: 78
Ratio: 5.026 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-12 x AA24514 ..

Align seg 1/1 to: AA24514 from: 1 to: 914

1 CACACCAGCAAAATCCCGCCCTCTGCTAGTTATGCAAAATATTCGCCA 50
|||||
600 AspThrSerLysPheProSerProLeuValValTyrAlaAsnIleArgG1 616
51 AGGACCTCCCAATCTCAGGGCCAGTGTACAGCCCTGATGATCAG 100
|||||
616 nGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerV 633

101 TGAATGAAAAACAGTTACCTTGGAACTACTGGATAATGAGCAGGTGCT 150
|||||
633 alasnGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAla 649
151 GATGCTACTAAGGATGACGGTGCTCTACTCAAGGTATTTTCACAACTATGCA 200
|||||
650 AspAlaThrLysAspAspGlyValTyrSerArgTyrPheThrThrTyrAs 666
201 CACGAATGGTAGATACAGTCTAAAGTGGGGCT 234
|||||
666 pThrAsnGlyArgTyrSerValLysValArgAla 677

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB73716

seq_documentation_block:
ID AAB73716 standard; Protein; 914 AA.

XX AAB73716;

XX 11-SEP-2001 (first entry)

XX Human CLCA1 protein, SEQ ID NO:2.

XX Human CLCA1; goblet cell; mouse Gob-5 orthologue; drug screening;
KW expression inhibition; antisense therapy; gene therapy;
KW chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic.

XX Homo sapiens.

XX Key Location/Qualifiers

XX FT Misc-difference 152

XX FT /note= "Encoded by AGG in AAH46124"

XX WO200138530-A1.

XX 31-MAY-2001.

XX 22-NOV-2000; 2000WO-JP08232.

XX 24-NOV-1999; 99JP-0333479.

XX 27-APR-2000; 2000JP-0127589.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Nakanishi A, Morita S;

XX WPI; 2001-355935/37.

XX N-PSDB; AAH46102, AAH46124.

XX New antisense nucleotide, useful for treatment and prevention of
PT bronchial asthma and chronic obstructive pulmonary disease

XX Claim 2; Page 76-80; 104pp; Japanese.

XX The invention relates to an antisense nucleotide targetted to the mouse
CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
CC the CLCA1 gene (coding sequence shown in AAH46102). The invention also
CC relates to an antibody specific for the Gob-5 protein, medical and
CC diagnostic compositions containing the antisense nucleotide or the
CC antibody, and methods and kits for screening for compounds which inhibit
CC the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.
CC The antisense oligonucleotides and antibody are therefore useful for the
CC treatment and prevention of bronchial asthma and chronic obstructive
CC pulmonary disease. The present sequence represents human CLCA1 protein.

XX Sequence 914 AA;

alignment_scores:

Quality: 392.00 Length: 78

Ratio: 5.026 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-12 x AAB73716

Align seg 1/1 to: AAB73716 from: 1 to: 914

1 GACACCAGCAAAATCCCGAGCCCTCTGCTAGTATTCGCAATATTCGCCA 50
|||||
600 AspThrSerLysPheProSerProLeuValTyrAlaAsnIleArgGI 616
51 AGGAGCCTCCCAATTCCTCAGGGCAGTGTCCACAGCCCTGATTGAATCAG 100
|||||
616 nGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerV 633
101 TGAATGAAAAACAGTTACCTTGGAACTACTGGATAATGAGCAGGTGCT 150
|||||
633 alasnGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAla 649
151 GATGCTACTAAGGATGACGGTGCTCTACTCAAGGTATTTTCACAACTATGCA 200
|||||
650 AspAlaThrLysAspAspGlyValTyrSerArgTyrPheThrThrTyrAs 666
201 CACGAATGGTAGATACAGTCTAAAGTGGGGCT 234
|||||
666 pThrAsnGlyArgTyrSerValLysValArgAla 677

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB74733

seq_documentation_block:

ID AAB74733 standard; Protein; 914 AA.

XX AAB74733;

XX 12-JUN-2001 (first entry)

XX Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.

XX Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
KW antiAlzheimer; antiparkinsonian; antimicrobial; vulnerary; gene therapy;
KW immune disorder; hyperproliferative disorder; cardiovascular disease;
KW cancer; angiogenic disorder; neurological disorder; infectious disease;
KW wound healing; regeneration; chemotaxis; chromosome 1.

XX Homo sapiens.

XX WO2000112775-A2.

XX 22-FEB-2001.

XX 16-AUG-2000; 2000WO-US22325.

XX 17-AUG-1999; 99US-0149182.

XX (HJMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ni J, Florence KA, Fiscella M, Wei P, Baker KP;

XX Birse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;

XX WPI; 2001-147550/15.

XX N-PSDB; AAF81787.

XX Nucleic acids encoding 25 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy

XX Claim 11; Page 459-460; 485pp; English.

XX AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733
CC to AAB74772. Human secreted proteins can have activities based on the
CC tissues and cells they are expressed in. Example of activities include:

CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;
CC antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;
CC vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic;
CC anticonvulsant; antiaizheimers; antiparkinsonian; antimicrobial; and
CC vulnerary. Human secreted proteins can be used in gene therapy and
CC vaccine. Human secreted protein nucleotide sequences (NAM1) and proteins
CC (PEP1) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. For example, NAM1
CC and PEP1 may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patients genome
CC that affect the activity of proteins by expressing inactive proteins or
CC to supplement the patients own production of polypeptides. Disorders that
CC may be prevented, diagnosed and/or treated include immune disorders,
CC hyperproliferative disorders (e.g. cancers), cardiovascular diseases,
CC angiogenic disorders, neurological disorders, infectious diseases and/or
CC for promoting wound healing, regeneration and /or chemotaxis. AAF81778 to
CC AAF81786 and AAB74732 represent sequences used in the exemplification of
CC the present invention.

XX Sequence 914 AA;

alignment_scores:
Quality: 392.00 Length: 78
Ratio: 5.026 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-12 x AAB74733 ..

Align seg 1/1 to: AAB74733 from: 1 to: 914

1 GACACCAGCAAAATCCCGAGCCCTCTGGTAGTTATGCAAAATATTCGCCA 50
|||||
600 AspThrSerLysPheProSerProLeuValValTyrAlaAsnIleAArgl 616
51 AGGAGCCCTCCCAATCTCAGGGCCAGTGTACAGCCCTGATTTGAATCAG 100
|||||
616 nGlyAlaSerProIleLeuA-GAlaSerValThrAlaLeuIleGluSerV 633
101 TGAATGGAAAAACAGTTACCTTGGAACTACTGGATAATTCGAGCAGGTGCT 150
|||||
633 alAsnGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAla 649
151 GATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTTCAACAATTATGA 200
|||||
650 AspAlaThrLysAspAspGlyValTyrSerArgTyrPheThrThrTyrAs 666
201 CACGAATGGTAGATACAGTGTAAAAGTGGGGCT 234
|||||
666 pThrAsnGlyArgTyrSerValLysValArgAla 677

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAG75474

seq_documentation_block:

ID AAG75474 standard; Protein: 925 AA.

XX AC AAG75474;

DT 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:6238.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 1.

XX Homo sapiens.

XX WO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX

PR 29-SEP-1999; 99US-0157137.

ER 03-NOV-1999; 99US-0163280.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX

DR WPI; 2001-235357/24.

XX

DR N-PSDB; AAB34879.

XX

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -

XX

PS Claim 11; Page 7686-7690; 9803pp; English.

XX

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB7789 represent sequences used in the exemplification of the
CC present invention.

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 925 AA;

alignment_scores:
Quality: 392.00 Length: 78
Ratio: 5.026 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-12 x AAG75474 ..

Align seg 1/1 to: AAG75474 from: 1 to: 925

1 GACACCAGCAAAATCCCGAGCCCTCTGGTAGTTATGCAAAATATTCGCCA 50
|||||
611 AspThrSerLysPheProSerProLeuValValTyrAlaAsnIleAArgl 627

51 AGGAGCCCTCCCAATCTCAGGGCCAGTGTACAGCCCTGATTTGAATCAG 100
|||||

627 nGlyAlaSerProIleLeuA-GAlaSerValThrAlaLeuIleGluSerV 644
|||||

101 TGAATGGAAAAACAGTTACCTTGGAACTACTGGATAATTCGAGCAGGTGCT 150
|||||

644 alAsnGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAla 660
|||||

151 GATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTTCAACAATTATGA 200
|||||

661 AspAlaThrLysAspAspGlyValTyrSerArgTyrPheThrThrTyrAs 677
|||||

201 CACGAATGGTAGATACAGTGTAAAAGTGGGGCT 234
|||||

677 pThrAsnGlyArgTyrSerValLysValArgAla 688
|||||

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74822

seq_documentation_block:

ID AAB74822 standard; Protein: 913 AA.

```

617  nGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerV 634
101  TGAATGGAAACAGTTACCTTGGAACTACTCGAATAATGCACGAGTGCT 150
634  aIaAsnGlyLysThrValThrLeuGluLeuLeuAsnGlyAlaGlyAla 650
151  GATGCTACTAAGGATGACGGGTGCTACTCAAGGTATTTCCACAACATTG 200
651  AspAlaThrLysAsnAspGlyValTyrSerArgPhePheThrAlaPheAs 667
201  CACGAATGTTAGATACAGTGTAAAGTCGGGGCT 234
667  pAlaAsnGlyArgTyrSerValLysIleIlePala 678

```

```
seq name: /STDS2/acadata/geneseq/geneseq/AA2001.DAT:AAB73715
```

seq documentation block:

seq_documentation_block	ID	AAR73715 standard	Protein: 913 AA

XX | AAR73715:

11-SEP-2007 (first entry)

XX DE Mouse Gob-5 protein. SEQ ID NO:1

PI Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;

PI Nicolaïdes NC, Zhou Y, Dong Q;

XX XX

DR WPI: 1999-550979/46.

DR N-PSDB; AAF81925.

[illegible]

PT New nucleic acid encoding calcium activated chloride channel, used to
PT identify, e.g. specific modulators for treating atopic allergy -
PT

xx
ps
ps
Claim 12: Fig 2: 75pp: English.

The present sequence represents the murine interleukin 9 (IL-9) induced calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins have anti-allergic, anti-asthmatic, anti-inflammatory and immunomodulatory activities. Compounds (A) that downregulate ICACC used to alleviate asthma (or more generally atopic allergy), while those (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and inflammatory bowel disease (IBD) (or other autoimmune diseases). Measuring levels of ICACC-1 can be used in the diagnosis of asthma (increased levels) or IBD (reduced levels), also for monitoring treatment of these conditions. The ICACC proteins can be used: (i) to raise specific antibodies (Ab). useful: (a) as immunoassay reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as (A) to disrupt bonding between ICACC and its ligands; and (iii) to identify modulators and binding partners. ICACC polynucleotides can be used to generate transgenic animals or recombinant cells, used to screen for antagonists, also as a source of therapeutic antisense agents or diagnostic probes (for quantifying mRNA expression, e.g. for identification of modulators).

AA	Sequence	913 AA;
SO		

alignment_scores:			
Quality:	339.00	Length:	78
Ratio:	4.403	Gaps:	0
Percent Similarity:	98.718	Percent Identity:	84.615

alignment block:

US-09-049-696-12 x AAB74822

Align seq 1/1 to: AAB74822 from: 1 to: 913

1 GACACCAGCAAAATTTCCCCAGCCCTCTGGTAGTTTATGCAAATAATTCCGCA 50
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
601 AsnThrGlVlvsPheProSerProValThrValTrAlaserIeArqGl 617

51 AGGAGCCTCCCCAATTCTCAGGGCCAGTGTACAGCCCTGATTGAATCAG 100

alignment_scores:			
Quality:	339.00	Length:	78
Ratio:	4.403	Gaps:	0
Percent Similarity:	98.718	Percent Identity:	84.615

alignment block:

Sequence 913 AA:

```
US-09-049-696-12 x AAB73715 ..
Align seg 1/1 to: AAB73715 from: 1 to: 913

1 GACACCAACAATCCCGAGCCCTGCTAGTCTTATGCAATATTCGCCA 50
601 AsnThrGlyLysPheProSerProValThrValThrAlaSerIleArgGI 617
51 AGGAGCCTCCCAATCTCAGGGCCAGTGTACAGCCCTGATTAATCAG 100
617 nGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerV 634
101 TGATGTAAGAACAGTTACCTTGGACTACTGGATAATGGAGCGGTGCT 150
634 alaSnGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAla 650
151 GATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCAACAACCTTATCA 200
651 AspalathrLysAsnAspGlyValThrSerArgPhePheThrAlaPheAs 667
201 CACGAATGGTAGACACTGTAAAGTCCGGGCT 234
667 pAlaAsnGlyArgThrValThrLeuLeuLeuValThrAlaPheAs 678

seq_name: /SDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA66749
seq_documentation_block:
ID AAY66749 standard; protein; 919 AA.
XX
AC AAY66749;
XX
DT 05-APR-2000 (first entry)
XX
DE Membrane-bound protein PRO1124.
XX
KW Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW pharmaceutical; receptor immunoadhesin; gene mapping.
XX
OS Homo sapiens.
XX
PN WO9963088-A2.
XX
PD 09-DEC-1999.
XX
PF 02-JUN-1999; 99WO-US1252.
XX
PR 02-JUN-1998; 98US-0087607.
PR 02-JUN-1998; 98US-0087609.
PR 02-JUN-1998; 98US-0087759.
PR 03-JUN-1998; 98US-0087827.
PR 04-JUN-1998; 98US-0088021.
PR 04-JUN-1998; 98US-0088025.
PR 04-JUN-1998; 98US-0088028.
PR 04-JUN-1998; 98US-0088029.
PR 04-JUN-1998; 98US-0088030.
PR 04-JUN-1998; 98US-0088033.
PR 04-JUN-1998; 98US-0088326.
PR 05-JUN-1998; 98US-0088167.
PR 05-JUN-1998; 98US-0088202.
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
```

```
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 07-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091678.
PR 07-JUL-1998; 98US-0091982.
PR 08-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
```

PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 20-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 26-AUG-1998; 98US-0097561.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 31-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 98US-0115565.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
DR WPI: 2000-072883/06.
DR N-PSDB; AAZ65055.
XX
PT Membrane-bound proteins and related nucleotide sequences
XX
PS claim 12; Fig 274; 822pp; English.
XX
CC The invention provides membrane-bound PRO polypeptides and
CC polynucleotides encoding them. The PRO sequences of the invention were
CC identified based on extracellular domain homology screening. The PRO
CC sequences have homology with proteins including LDL receptors, TIE
CC ligands and various enzymes. The membrane-bound proteins and receptor
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
CC immunoadhesins, for instance, can be used as therapeutic agents to block
CC receptor-ligand interactions. The membrane-bound proteins can also be
CC employed for screening of potential peptide or small molecule inhibitors
CC of the relevant receptor/ligand interaction. The PRO encoding sequences
CC are useful as hybridization probes, in chromosome and gene mapping and in
CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC will also be useful for the preparation of PRO polypeptides, especially
CC by recombinant techniques.
XX
SQ Sequence 919 AA;

alignment_scores:
Quality: 277.00 Length: 78
Ratio: 4.074 Gaps: 0
Percent Similarity: 87.179 Percent Identity: 69.231
alignment_block:

US-09-049-696-12 x AAY66749 ..
Align seg 1/1 to: AAY66749 from: 1 to: 919
1 GACACCAGCAAAATCCCCAGCCCTCTGTTAGTTATGCAAAATATTCGCCA 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
603 AspValAsnSerPheProSerPrometIleValTyrAlaGluIleLeuG1 619
51 AGGAGCCTCCCAATTTCTCAGGCCAGTGTACAGCCCTGATTGAATCAG 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
619 nGlyTyrValProValLeuGlyAlaAsnValThrAlaPheIleGluSerG 636
101.TGAATGGAACACAGTTACTTGGAACTACTGGATAATGGAGCAGTGCT 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
636 InAsnGlyHisThrGluValLeuGluLeuLeuAspAsnGlyAlaGlyAla 652
151 GATGCTACTAAGGATGACGCTGTCTACTCAAGGTATTTTCACAACTTATCA 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
653 AspSerPheLysAsnAspGlyValTyrSerArgTyrPheThrAlaTyrTh 669
201 CACGAATGCTAGATACAGTGTAAAGTGGCGGCT 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||
669 rGluAsnGlyArgTyrSerLeuLysValArgAla 680
seq_name: /SIDS2/qcndata/geneseq/geneseq/AA2001.DAT: AAB87560
seq_documentation_block:
ID AA387560 standard; Protein; 919 AA.
XX
AC AAB87560;
XX
DT 15-MAY-2001 (first entry)
XX
DE Human PRO1124.
XX
KW Human; PRO protein; mapping.
XX
OS Homo sapiens.
XX
PN WC200116318-A2.
XX
PD 08-MAR-2001.
XX
PF 24-AUG-2000; 2000WO-US23328.
XX
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 07-DEC-1999; 99US-0169495.
PR 05-DEC-1999; 99US-0170262.
PR 11-JAN-2000; 2000US-0175481.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 01-MAR-2000; 2000WO-US05601.
PR 03-MAR-2000; 2000US-0187202.
PR 25-APR-2000; 2000US-019397.
PR 22-MAY-2000; 2000WO-US14042.
PR 05-JUN-2000; 2000US-0209832.
XX
PA (GETH) GENENTECH INC.
XX
PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
XX
DR WPI: 2001-183260/18.
DR N-PSDB; AAF92092.
XX
PT Eighty four nucleic acids encoding PRO polypeptides, useful in
PT molecular biology, including use as hybridization probes, and in
PT chromosome and gene mapping.
XX
PS Claim 12; Fig 70; 278pp; English.
XX

CC The present sequence is a human PRO polypeptide (secreted and
 CC transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
 CC anti-PRO antibodies are useful for preparation of a medicament useful in
 CC the treatment of a condition which is responsive to the PRO protein,
 CC agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
 CC employed as molecular weight markers for protein electrophoresis. The PRO
 CC coding sequence has applications in molecular biology, including use as
 CC hybridisation probes, and in chromosome and gene mapping.
 XX
 SQ Sequence 919 AA;

alignment_scores:
 Quality: 277.00 Length: 78
 Ratio: 4.074 Gaps: 0
 Percent Similarity: 87.179 Percent Identity: 69.231

alignment_block:

US-09-049-696-12 x AAB87560 ..

Align seg 1/1 to: AAB87560 from: 1 to: 919

1 GACACAGCAATCCCGCCCTCTGCTAGTTTATGCAATATTCGCCA 50
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 603 AspValAsnSerPheProSerProMetileValTyrAlaGluileLeuG1 619
 51 AGGAGCCTCCCAATCTCAGGGCCAGTGTACAGCCCTGATTGAATCAG 100
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 619 nGlyTyrValProValLeuGlyAlaAsnValThrAlaPheileGluSerG 636
 101 TGAATGGAAACACAGTACCTTGGAACTACTGGATGATGGAGCAGGTGCT 150
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 636 InAsnGlyHisThrGluValLeuGluLeuLeuAspAsnGlyAlaGlyAla 652
 151 GATGCTACTAAGGATGACGGTCTCTACTCAAGGTATTTCAACACTTATGA 200
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 653 AspSerPheLysAsnAspGlyValTyrSerArgTyrPheThrAlaTyrTh 669
 201 CACGAATGGTAGATACACTGTAAAGTCGGGCT 234
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 669 rGluAsnGlyArgTyrSerLeuLysValArgAla 680

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB65272

seq_documentation_block:

ID AAB65272 standard; Protein; 919 AA.

XX AAB65272:

DT 02-APR-2001 (first entry)

DE Human PRO1124 (UNQ562) protein sequence SEQ ID NO:379.

XX Human; secreted and transmembrane protein; PRO; cytostatic;
 KW cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
 KW diagnostic assay.

XX Homo sapiens.

XX WO200073454-A1.

XX 07-DEC-2000.

XX 30-MAR-2000; 2000WO-US08439.

XX 02-JUN-1999; 99WO-US12252.

XX 23-JUN-1999; 99US-0141037.

XX 07-JUL-1999; 99US-0143048.

XX 20-JUL-1999; 99US-0144758.

XX 26-JUL-1999; 99US-0145698.

XX 28-JUL-1999; 99US-0146222.

XX 17-AUG-1999; 99US-0149396.

XX 15-SEP-1999; 99WO-US21090.

PR 15-SEP-1999; 99WO-US21547.
 PR 08-OCT-1999; 99US-0158663.
 PR 30-NOV-1999; 99WO-US28313.
 PR 01-DEC-1999; 99WO-US28301.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00376.
 PR 11-FEB-2000; 2000WO-US03565.
 PR 18-FEB-2000; 2000WO-US04341.
 PR 22-FEB-2000; 2000WO-US04414.
 PR 24-FEB-2000; 2000WO-US04914.
 PR 24-FEB-2000; 2000WO-US05004.
 PR 02-MAR-2000; 2000WO-US05841.
 PR 15-MAR-2000; 2000WO-US06884.
 PR 20-MAR-2000; 2000WO-US07377.
 XX
 PA (GETH) GENENTECH INC.

PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;

PI Feirara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
 PI Grimaldi CJ, Gurney AL, Kijavirij, Napier MA, Pan J, Paoni NF;
 PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
 PI Zhang Z;

XX WPI: 2001-032160/04.

DR N-PSDB; AAF44241.

XX
 PT PRO polynucleotides used to produce polypeptides used to target
 PT bioactive molecules such as toxins, radiolabels or antibodies, to
 PT specific cells, to cause targeted cell death -

XX Claim 12: Fig 274; 935pp; English.

XX The present invention describes human secreted and transmembrane PRO
 CC proteins. The PRO proteins have cytostatic activity. The PRO proteins
 CC can be used for targeted delivery of bioactive molecules, such as
 CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
 CC sequences, and their fragments, can be used as hybridisation probes, in
 CC chromosomal and gene mapping, and in the generation of anti-sense RNA
 CC and DNA. They may also be used to produce transgenic animals which are
 CC used to develop and screen therapeutically useful reagents. The PRO
 CC nucleotide and protein sequence can be used for tissue typing and in
 CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
 CC AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
 CC in the isolation of human PRO sequences. AAF44087 to AAF44269 and
 CC AAB65154 to AAB65300 represent human PRO polynucleotide and protein
 CC sequences given in the exemplification of the present invention.

XX Sequence 919 AA;

alignment_scores:

Quality: 277.00 Length: 78
 Ratio: 4.074 Gaps: 0
 Percent Similarity: 87.179 Percent Identity: 69.231

alignment_block:

US-09-049-696-12 x AAB65272 ..

Align seg 1/1 to: AAB65272 from: 1 to: 919

1 GACACAGCAATCCCGCCCTCTGCTAGTTTATGCAATATTCGCCA 50
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 603 AspValAsnSerPheProSerProMetileValTyrAlaGluileLeuG1 619
 51 AGGAGCCTCCCAATCTCAGGGCCAGTGTACAGCCCTGATTGAATCAG 100
 |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 619 nGlyTyrValProValLeuGlyAlaAsnValThrAlaPheileGluSerG 636
 101 TGAATGGAAACACAGTACCTTGGAACTACTGGATGATGGAGCAGGTGCT 150
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 636 InAsnGlyHisThrGluValLeuGluLeuLeuAspAsnGlyAlaGlyAla 652

614 AspSerLeuHisPheProHisProValMetIleTyrAlaAsnValLysG1 630

FT	..	Modified-site	/note= 2	"Phosphorylation site"
FT	FT	Modified-site	/note= 76	"Phosphorylation site"
FT	FT	Modified-site	/note= 92	"Phosphorylation site"
FT	FT	Modified-site	/note= 270	"Phosphorylation site"
FT	FT	Modified-site	/note= 318	"Phosphorylation site"
FT	FT	Modified-site	/note= 384	"Phosphorylation site"
FT	FT	Modified-site	/note= 557	"Phosphorylation site"
FT	FT	Modified-site	/note= 568	"Phosphorylation site"
FT	FT	Modified-site	/note= 593	"Phosphorylation site"
FT	FT	Modified-site	/note= 652	"Phosphorylation site"
FT	FT	Modified-site	/note= 685	"Phosphorylation site"
FT	FT	Modified-site	/note= 822	"Phosphorylation site"
FT	FT	Modified-site	/note= 823	"Phosphorylation site"
FT	FT	Modified-site	/note= 858	"Phosphorylation site"
FT	FT	Modified-site	/note= 927	"Phosphorylation site"
FT	FT	Modified-site	/note= 74	"Phosphorylation site"
FT	FT	Modified-site	/note= 97	"Glycosylation site"
FT	FT	Modified-site	/note= 150	"Glycosylation site"
FT	FT	Modified-site	/note= 231	"Glycosylation site"
FT	FT	Modified-site	/note= 235	"Glycosylation site"
FT	FT	Modified-site	/note= 253	"Glycosylation site"
FT	FT	Modified-site	/note= 291	"Glycosylation site"
FT	FT	Modified-site	/note= 521	"Glycosylation site"
FT	FT	Modified-site	/note= 555	"Glycosylation site"
FT	FT	Modified-site	/note= 579	"Glycosylation site"
FT	FT	Modified-site	/note= 636	"Glycosylation site"
FT	FT	Modified-site	/note= 821	"Glycosylation site"
FT	FT	Modified-site	/note= 937	"Glycosylation site"
FT	FT	Peptide	/note= 1..31	"Glycosylation site"
FT	FT	Domain	/label= Signal_peptide 900..926	
FT	FT		/label= Transmembrane_domain	
XX	XX	WC200012711-A2.		
PD	09- MAR- 2000.			
XX	02- SEP- 1999;		99WO-US20468.	
XX				
PR	02- SEP- 1998;		98US-0145815.	
PR	12- NOV- 1998;		98US-0191283.	
PR	09- DEC- 1998;		98US-0208821.	
PR	26- JAN- 1999;		99US-0237506.	
PR	10- FEB- 1999;		99US-0247891.	

OM of: US-09-049-696-11 to: SPTREMBL_17:* out_format : pfs
Date: Mar 30, 2002 2:46 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+g2p.model -DEV=xlp  
-Q/-cgn2_1/USPTO_spo1/US09049696/runat_28032002_145238_2053/app_query.fasta_1.12579  
-DB=SPTREMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -FGAPEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -MINLEN=0 -MAXLEN=200000000  
-USER=US09049696 -CGN1_1_705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY  
-WAIT -THREADS=1
```

Search information block:

Query: US-09-049-696-11
Query length: 191
Database: SPTREMBL_17:*
Database sequences: 473505
Database length: 14627329
Search time (sec): 805.760000

score_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
sp_human:095151	+	315.00	708.53	3.7e-32	914 ! O95151 homo sapiens (human). ca
sp_human:09UPC6	+	315.00	708.53	3.7e-32	914 ! Q9UPC6 homo sapiens (human). ca
sp_human:09UNF6	+	315.00	708.53	3.7e-32	914 ! Q9UNF6 homo sapiens (human). ca
sp_mammal:09TRB5	+	248.00	552.78	1.8e-23	917 ! Q9TRB5 sus scrofa (pig). epithe
sp_rodent:088826	+	246.00	548.17	3.2e-23	913 ! O88826 mus musculus (mouse). gc
sp_rodent:09D726	+	246.00	548.17	3.2e-23	913 ! Q9D726 mus musculus (mouse). ch
sp_human:09UNX1	+	207.00	462.63	3.6e-18	469 ! Q9UNX1 homo sapiens (human). cd
sp_human:09UNF7	+	207.00	457.49	3.6e-18	917 ! Q9UNF7 homo sapiens (human). ca
sp_mammal:018742	+	153.00	333.08	3.5e-11	794 ! O18742 bos taurus (bovine). lu-
sp_mammal:018743	+	153.00	332.84	3.5e-11	820 ! O18743 bos taurus (bovine). lu-
sp_mammal:018741	+	153.00	332.08	3.5e-11	905 ! O18741 bos taurus (bovine). lu-
sp_rodent:088860	+	130.00	278.66	3.3e-08	901 ! O88860 mus musculus (mouse). ch
sp_rodent:09QX15	+	130.00	278.65	3.3e-08	902 ! Q9QX15 mus musculus (mouse). ca
sp_rodent:09R070	+	128.50	275.16	5.2e-08	902 ! Q9R070 mus musculus (mouse). ca
sp_rodent:09E0R4	+	128.50	275.16	5.2e-08	902 ! Q9E0R4 mus musculus (mouse). en
sp_human:09UOC9	+	124.50	265.52	1.7e-07	943 ! Q9UOC9 homo sapiens (human). ch
sp_human:09Y6N2	+	124.50	265.52	1.7e-07	943 ! Q9Y6N2 homo sapiens (human). ch
sp_bacteria:09PKZ0	+	70.00	145.48	1.99	397 ! Q9PKZ0 chlamydia muridarum. hyp
sp_human:09BV32	+	67.00	136.10	4.83	544 ! Q9BV32 homo sapiens (human). si
sp_human:094929	+	67.00	134.35	4.81	683 ! O94929 homo sapiens (human). ki
sp_vertebrate:013275	-	66.50	148.32	5.79	90 ! O13275 clupea pallasii (pacific
sp_vertebrate:013274	-	66.50	148.32	5.78	94 ! O13274 clupea pallasii (pacific
sp_plant:09S2C6	+	66.50	136.92	5.64	420 ! Q9S2C6 arabidopsis thaliana (md
sp_plant:09LK15	-	66.50	135.61	5.62	498 ! Q9LK15 arabidopsis thaliana (md
sp_plant:09SOH9	-	66.50	135.29	5.61	519 ! Q9SOH9 arabidopsis thaliana (md
sp_plant:09FUH7	-	66.50	135.29	5.61	519 ! Q9FUH7 brassica napus (rape). h
sp_human:09NNW7	+	66.00	130.02	6.46	888 ! Q9NNW7 homo sapiens (human). co
sp_invertebrate:09VLT6	-	65.00	118.84	8.53	2820 ! Q9VLT6 drosophila melanogast
sp_plant:09LK16	+	63.50	128.67	13.74	496 ! Q9LK16 arabidopsis thaliana (md
sp_fungi:09USG5	+	63.50	120.56	13.50	1429 ! Q9USG5 schizosaccharomyces pom
sp_mhc:002870	+	63.00	130.54	16.06	334 ! O02870 gallus gallus (chicken).
sp_mhc:002869	+	63.00	130.18	16.05	350 ! O02869 gallus gallus (chicken).
sp_bacteria:09CFT0	+	63.00	129.69	16.03	373 ! Q9CFT0 lactococcus lactis (subs
sp_phase:09A2R7	+	63.00	129.69	16.03	373 ! Q9A2R7 bacteriophage bil286. or
sp_virus:036401	+	63.00	125.06	15.86	683 ! O36401 acicaphine herpesvirus
sp_bacteria:069385	+	63.00	119.61	15.67	1390 ! O69385 streptococcus mutans. g
sp_bacteria:069382	+	63.00	119.26	15.66	1455 ! O69382 streptococcus mutans. g
sp_bacteria:069391	+	63.00	119.26	15.66	1455 ! O69391 streptococcus mutans. g
sp_bacteria:069397	+	63.00	119.26	15.66	1455 ! O69397 streptococcus mutans. g
sp_fungi:001759	+	62.50	128.27	18.59	386 ! Q01759 pneumocystis carinii. an

sp_fungi:013934 + 62.50 127.55 18.56 424 ! O13934 schizosaccharomyces p
sp_virus:039990 + 62.50 123.81 18.41 690 ! O39990 human herpesvirus 7.
sp_plant:09LVF2 + 62.50 121.32 18.31 955 ! Q9LVF2 arabidopsis thaliana
sp_fungi:001805 + 62.50 120.42 18.27 1075 ! Q01805 pneumocystis carinii
sp_fungi:001830 + 62.50 120.41 18.27 1076 ! Q01830 pneumocystis carinii
seq_name: sp_human:095151
seq_documentation_block:
ID O95151 PRELIMINARY; PRT; 914 AA.
AC O95151:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CALCITUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D.; Eible R.C.; Ji H.L.; Schreuer K.D.; Fuller C.M.;
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
of human CLCAL, the first human member of the family of Ca2+-activated
Cl-channel proteins.";
RL Genomics 54:200-214(1998).
DR EMBL; AF039400; AAC95428.1; -;
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; P500153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; P50234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 914 AA; 100174 MW; 7066440C46557FA3 CRC64;

alignment_scores:
! Quality: 315.00 Length: 63
! Ratio: 5.000 Gaps: 0
! Percent Similarity: 100.00 Percent Identity: 100.00

alignment_block:
US-09-049-696-11 x O95151 ..
Align seg 1/1 to: O95151 from: 1 to: 914
! ! GGCCTTCTAGTGGACAAAAACACCAAAATGCCTTACCTCCAAATCCCAAG 50
541 GlypheValValAspLysAsnThrLysMetAlaTyrLeuGlnIleProG1 557
51 CATTGCTAAGTTGGCACTGGGAATACAGTCTCAAGCAAGCTCACAAA 100
557 ylleAlaLysValGlyThrIrpLysTyrSerLeuGlnAlaSerSerGlnT 574
101 CCTGACCTGACTGTGCTCCGCTCCGCTCAATGCTACCTGCCTCCCA 150
574 hrLeuThrLeuThrValThrSerArgAlaSerAsnAlaThrLeuProPro 590

151 ATTACCTGACTTCCAAAACAGAACAGACACCAAGCAAGCAAGCAAA 189
591 IleThrValThrSerLysThrAsnLysAspThrSerLys 603

seq_name: sp_human:09UPC6

seq_documentation_block:

ID O9UPC6 PRELIMINARY; PRT; 914 AA.
AC O9UPC6:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

```
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Elble R.C., Ji H.L., Schreier K.D., Fuller C.M.,
RA Pauli B.U.
RA "Genomic cloning, molecular characterization, and functional analysis
RT of human CLCA1, the first human member of the family of Ca2+-activated
RT Cl-channel proteins."
RL Genomics 54:200-214(1998).
DR EMBL; AF039401; AAC95429.1; -.
DR InterPro; IPR00131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR SEQUENCE 914 AA; 100160 MW; 7761362B66752AA3 CRC64;
SQ
alignment_scores:
  Quality: 315.00 Length: 63
  Ratio: 5.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-049-696-11 x Q9UPC6
..
Align seg 1/1 to: Q9UPC6 from: 1 to: 914
1 GGCTTTGTAGTGACAAAACCAAAATGGCTACCTCCAAATCCAGG 50
|||||
541 GlyPheValValAspLysAsnThrLysMetAlaTyrLeuGlnIleProG1 557
51 CATTGCTAAGTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAAA 100
|||||
557 ylleAlaLysValGlyThrTrpLysTyrSerLeuGlnAlaSerSerGlnT 574
101 CCTTGACCTGACTGTCACGTCCTGGCTGCTCAATGCTACCTGCTCCA 150
|||||
574 hrLeuThrLeuThrValThrSerArgAlaSerAsnAlaThrLeuProPro 590
151 ATTACAGTCACTTCCAAAACGAACAGCAAGCACACGACAAA 189
|||||
591 IleThrValThrSerLysThrAsnLysAspThrSerLys 603
seq_name: sp_human:Q9UNF6
seq_documentation_block:
ID Q9UNF6 PRELIMINARY; PRT; 914 AA.
AC Q9UNF6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN.
GN CACCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=SMALL INTESTINE, AND COLON;
RX MEDLINE=99364503; PubMed=10437792;
RA Agnel M., Vermet T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
RT chloride channel (CACCL) family predominantly expressed in the
RT digestive tract and trachea."
RL FEBS Lett. 455:295-301(1999).
```

```
DR EMBL; AF127036; AAD25487.1; -.
DR InterPro; IPR00131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
DR SEQUENCE 914 AA; 100202 MW; B58B451D97887D6B CRC64;
SQ
alignment_scores:
  Quality: 315.00 Length: 63
  Ratio: 5.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-049-696-11 x Q9UNF6
..
Align seg 1/1 to: Q9UNF6 from: 1 to: 914
1 GGCTTTGTAGTGACAAAACCAAAATGGCTACCTCCAAATCCAGG 50
|||||
541 GlyPheValValAspLysAsnThrLysMetAlaTyrLeuGlnIleProG1 557
51 CATTGCTAAGTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAAA 100
|||||
557 ylleAlaLysValGlyThrTrpLysTyrSerLeuGlnAlaSerSerGlnT 574
101 CCTTGACCTGACTGTCACGTCCTGGCTGCTCAATGCTACCTGCTCCA 150
|||||
574 hrLeuThrLeuThrValThrSerArgAlaSerAsnAlaThrLeuProPro 590
151 ATTACAGTCACTTCCAAAACGAACAGCAAGCACACGACAAA 189
|||||
591 IleThrValThrSerLysThrAsnLysAspThrSerLys 603
seq_name: sp_mammal:Q9TUB5
seq_documentation_block:
ID Q9TUB5 PRELIMINARY; PRT; 917 AA.
AC Q9TUB5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN.
GN AECC.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=ILEAL MUCOSA;
RA Gaspar K.J., Gabriel S.E., Racette K.J., Forsyth G.W.;
RT "A CDNA involved in porcine exocrine chloride conductance."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF095584; AAF00077.1; -.
DR InterPro; IPR00131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR SEQUENCE 917 AA; 100735 MW; 649F41944F7EAD19 CRC64;
SQ
alignment_scores:
  Quality: 248.00 Length: 62
  Ratio: 4.351 Gaps: 0
Percent Similarity: 91.935 Percent Identity: 77.419
alignment_block:
US-09-049-696-11 x Q9TUB5
..
Align seg 1/1 to: Q9TUB5 from: 1 to: 917
```

```

Align seg 1/1 to: Q9D/26 from: 1 to: 913

1 GGCTTTGTAGTGGCAAAAAACCAAAATGGCTACCTCCAATCCAGG 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
542 GlyPheilleLeuaspThrThrLysValAlaTyrLeuGlnValProG1 558
51 CATTGCTAAGTTGGCACTTGGAAATACAGCTGCAAGCAAGCTCAAA 1000
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
558 yThrAlaLysValGlyPheThrLysTyrSerIleGlnAlaSerSerGlnT 575
101 CCTTGACCCGTACGTGCACGTCGCGTGCATGCTACCTCGCCTCCA 1500
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
575 hrLeuThrLeuThrValThrSerArgAlaAlaSerAlaThrLeuProPro 591
151 ATTTACAGTGACTTCCAAAACCAAGGACACCCAGCAAA 189
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```
592 IleThrValThrProValValAsnLysAsnThrGlyLys 604
seq_name: sp_human:Q9NXP1
seq_documentation_block:
ID Q9NXP1 PRELIMINARY; PRT; 469 AA.
AC Q9NXP1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CDNA FLJ20131 FIS, CLONE COL06357.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000138; BAA0969.1;
SQ SEQUENCE 469 AA; 50694 MW; 6765986C886B3E36 CRC64;

alignment_scores:
Quality: 207.00 Length: 63
Ratio: 3.632 Gaps: 1
Percent Similarity: 90.476 Percent Identity: 63.492

alignment_block:
US-09-049-696-11 x Q9NXP1
Align seg 1/1 to: Q9NXP1 from: 1 to: 469
4 TTTGTAGTGACAAAACACCAAAATGGCTACCTCCAAATCCAGGCAT 53
||||| : : : : : : : : : : : : : : : : : : : :
95 PhetrValAspAlaThrSerLysMetAlaTyLeuSerIleProGlyTh 111
||||| : : : : : : : : : : : : : : : : : : : :
54 TGCTAAGTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCA.....C 97
||||| : : : : : : : : : : : : : : : : : : : :
111 rAlaLysValGlyThrTrpAlaTyAsnLeuGlnAlaLysAlaAsnProG 128
||||| : : : : : : : : : : : : : : : : : : : :
98 AAACCTTGACCTGACGTGCACGTCGCGTCCGTCACATGCTACCTGCCT 147
||||| : : : : : : : : : : : : : : : : : : : :
128 luThrLeuThrIleThrValThrSerArgAlaAlaAsnSerSerValPro 144
||||| : : : : : : : : : : : : : : : : : : : :
148 CCAATTACAGTACTTCCAAACCAAGCAAGCAGCAGC 186
||||| : : : : : : : : : : : : : : : : : : : :
145 ProIleThrValAsnAlaLysMetAsnLysAspValAsn 157

seq_name: sp_human:Q9UNF7
seq_documentation_block:
ID Q9UNF7 PRELIMINARY; PRT; 917 AA.
AC Q9UNF7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 2.
GN CAC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Agnel M., Vermet T., Culouscou J.M.;
RL "Identification of three novel members of the calcium-dependent
```

```
RT chloride channel (CaCC) family predominantly expressed in the
RT digestive tract and trachea.";
RL FEBS Lett 455:295-301(1999).
DR EMBL; AF127035; AAD48398.1; -.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00092; vwa; 1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 917 AA; 101153 MW; EA01090D781BEB95 CRC64;

alignment_scores:
Quality: 207.00 Length: 63
Ratio: 3.632 Gaps: 1
Percent Similarity: 90.476 Percent Identity: 63.492

alignment_block:
US-09-049-696-11 x Q9UNF7
Align seg 1/1 to: Q9UNF7 from: 1 to: 917
4 TTTGTAGTGACAAAACACCAAAATGGCTACCTCCAAATCCAGGCAT 53
||||| : : : : : : : : : : : : : : : : : : : :
543 PhetrValAspAlaThrSerLysMetAlaTyLeuSerIleProGlyTh 559
||||| : : : : : : : : : : : : : : : : : : : :
54 TGCTAAGTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCA.....C 97
||||| : : : : : : : : : : : : : : : : : : : :
559 rAlaLysValGlyThrTrpAlaTyAsnLeuGlnAlaLysAlaAsnProG 576
||||| : : : : : : : : : : : : : : : : : : : :
98 AAACCTTGACCTGACGTGCACGTCGCGTCCGTCACATGCTACCTGCCT 147
||||| : : : : : : : : : : : : : : : : : : : :
576 luThrLeuThrIleThrValThrSerArgAlaAlaAsnSerSerValPro 592
||||| : : : : : : : : : : : : : : : : : : : :
148 CCAATTACAGTACTTCCAAACCAAGCAAGCAGCAGC 186
||||| : : : : : : : : : : : : : : : : : : : :
593 ProIleThrValAsnAlaLysMetAsnLysAspValAsn 605

seq_name: sp_mammal:O18742
seq_documentation_block:
ID O18742 PRELIMINARY; PRT; 794 AA.
AC O18742;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001262; AAB86530.1; -.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 794 AA; 88509 MW; B695E7256FC2C632 CRC64;

alignment_scores:
Quality: 153.00 Length: 65
Ratio: 3.188 Gaps: 2
Percent Similarity: 73.846 Percent Identity: 52.308

alignment_block:
US-09-049-696-11 x O18742
Align seg 1/1 to: O18742 from: 1 to: 794
```

seq documentation block:

```

DR MGD; MGI:1316732; Clcal.
DR InterPro: IPR002035; vWFA.
DR SMART; SM00327; vWA; 1.
DR PROSITE; PS50234; vWFA; 1.
SQ SEQUENCE 901 AA; 100038 MW; 13DF068BF3337AAB CRC64;

alignment_scores:
  Quality: 130.00      Length: 64
  Ratio: 2.766        Gaps: 2
  Percent Similarity: 73.438      Percent Identity: 43.750

alignment_block:
US-09-049-696-11 x 088860 ..
Align seg 1/1 to: 088860 from: 1 to: 901

4 TTTGTAGTGGACAAA...AACACCAAAATGGCTACCTCCAAATCCAGG 50
||| ||||| ||| ||| ||||| |||||
545 PheGlnAspAspLysLeuAsnIleArgSerAlaArgLeuGlnIleProG1 561
| ||||| ||||| ||||| ||||| ||||| |||||
51 CATTGCTAAGTTGGCAGCTTGGAAATACAGTCTGCAAGCAAGC...TCAC 97
| ||||| ||||| ||||| ||||| ||||| |||||
561 yThrAlaGluThrGlyThrTrpThrTyrSerIleThrGlyThrLysSerG 578
| ||||| ||||| ||||| ||||| ||||| |||||
98 AAACCTTGACCTGACTGTCCAGTCCTCCGTCCTCAATGCTACCTGCCT 147
||| ||||| ||||| ||||| ||||| |||||
578 lnLeuIleThrMetThrValThrArgAlaArgSerProThrMetGlu 594
| ||||| ||||| ||||| ||||| ||||| |||||
148 CCAATTACAGTACTTCCAAAGCAAGACAGCAGCAGCAAA 189
||| ||||| ||||| ||||| ||||| |||||
595 ProLeuAlaThrAlaHisMetSerGlnSerThrAlaGln 608

seq_name: sp_rodent:Q9QX15

seq_documentation_block:
ID Q9QX15 PRELIMINARY; PRT; 902 AA.
AC Q9QX15;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE CA(2+)-SENSITIVE CHLORIDE CHANNEL 2.
GN CLCA2 OR CLCAL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=WHOLE LUNG;
MEDLINE=99041980; PubMed=9822685;
RA Gandhi R., Elble R.C., Gruber A.D., Schreur K.D., Ji H.-L.,
RA Fuller C.M., Pauli B.U.;
RT "Molecular and functional characterization of a calcium-sensitive
chloride channel from mouse lung."
RL J. Biol. Chem. 273:32096-32101(1998).
DR EMBL; AF047838; AAC79982.1; -.
DR MGD; MGI:1316732; Clcal.
DR InterPro: IPR002035; vWFA.
DR SMART; SM00327; vWA; 1.
DR PROSITE; PS50234; vWFA; 1.
SQ SEQUENCE 902 AA; 100107 MW; D9A915E698EA2D85 CRC64;

alignment_scores:
  Quality: 130.00      Length: 64
  Ratio: 2.766        Gaps: 2
  Percent Similarity: 73.438      Percent Identity: 43.750

alignment_block:
US-09-049-696-11 x Q9QX15 ..
Align seg 1/1 to: Q9QX15 from: 1 to: 902

```

```

4 TTTGTAGTGGACAAA...AACACCAAAATGGCTACCTCCAAATCCAGG 50
||| ||||| ||| ||| ||||| |||||
545 PheGlnAspAspLysLeuAsnIleArgSerAlaArgLeuGlnIleProG1 561
| ||||| ||||| ||||| ||||| ||||| |||||
51 CATTGCTAAGTTGGCAGCTTGGAAATACAGTCTGCAAGCAAGC...TCAC 97
| ||||| ||||| ||||| ||||| ||||| |||||
561 yThrAlaGluThrGlyThrTrpThrTyrSerIleThrGlyThrLysSerG 578
| ||||| ||||| ||||| ||||| ||||| |||||
98 AAACCTTGACCTGACTGTCCAGTCCTCCGTCCTCAATGCTACCTGCCT 147
||| ||||| ||||| ||||| ||||| |||||
578 lnLeuIleThrMetThrValThrArgAlaArgSerProThrMetGlu 594
| ||||| ||||| ||||| ||||| ||||| |||||
148 CCAATTACAGTACTTCCAAAGCAAGACAGCAGCAGCAAA 189
||| ||||| ||||| ||||| ||||| |||||
595 ProLeuAlaThrAlaHisMetSerGlnSerThrAlaGln 608

seq_name: sp_rodent:Q9R070

seq_documentation_block:
ID Q9R070 PRELIMINARY; PRT; 902 AA.
AC Q9R070;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE CA(2+)-SENSITIVE CHLORIDE CHANNEL 2.
GN CLCA2 OR CACC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY;
MEDLINE=20012773; PubMed=10544033;
RA Lee D., Ha S., Kho Y., Kim J., Cho K., Baik M., Choi Y.;
RT "Induction of mouse Ca(2+)-sensitive chloride channel 2 gene during
involvement of mammary gland."
RL Biochem. Biophys. Res. Commun. 264:933-937(1999).
DR EMBL; AF108501; AAF12731.1; -.
DR MGD; MGI:1931471; Clca2.
DR InterPro: IPR002035; vWFA.
DR SMART; SM00327; vWA; 1.
DR PROSITE; PS50234; vWFA; 1.
SQ SEQUENCE 902 AA; 99866 MW; 0FB746CF5FB3D07F CRC64;

alignment_scores:
  Quality: 128.50      Length: 58
  Ratio: 2.920        Gaps: 1
  Percent Similarity: 75.862      Percent Identity: 43.103

alignment_block:
US-09-049-696-11 x Q9R070 ..
Align seg 1/1 to: Q9R070 from: 1 to: 902

19 AACACCAAAATGGCTACCTCCAAATCCAGGCTTGGTAAAGTTGGCAGC 68
||| ||||| ||| ||| ||||| |||||
551 AsnIleArgSerAlaArgLeuGlnIleProGlyThrAlaGluThrGlyTh 567
| ||||| ||||| ||||| ||||| ||||| |||||
69 TTGGAATACAGTCTGCAAGCAAGC...TCACAAACCTTGACCTGACTG 115
||| ||||| ||||| ||||| ||||| |||||
567 rTrpThrTyrSerIleThrGlyThrLysSerGlnLeuIleThrMetThrV 584
| ||||| ||||| ||||| ||||| ||||| |||||
116 TCACGTCCCGTGGTCCCAATGCTACCTCCCTCCCAATACAGTACTTCC 165
||| ||||| ||||| ||||| ||||| |||||
584 alThrThrArgAlaArgSerProThrMetGluProLeuAlaThrAla 600
| ||||| ||||| ||||| ||||| ||||| |||||
166 AAAACGAAACAGCAGCAGCAGCAAA 189
||| ||||| ||||| ||||| ||||| |||||
601 HisMetSerGlnSerThrAlaGln 608

```


This Page Blank (uspto)

OM of: US-09-049-696-11 to: SwissProt_39:* out_format : pfs

Date: Mar 30, 2002 2:52 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+2p.model -DEV=xlp
-O=/cgn2.1/USPFO_spool/US09049696/runat_28032002_145238_2085/app_query.fasta_1.12579
-DB=SwissProt_39 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DLOCALIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 @CGN1_1_165 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-11
Query length: 191
Database: SwissProt_39:*
Database sequences: 100059
Database length: 36664827
Search time (sec): 306.030000

score_list:

Sequence	Strd Orig	Zscore	EScore	len	Documentation
SwissProt_39:ECIC_BOVIN	+	146.00	310.04	1.9e-10	903 ! P54281 bos taurus (bovine).
SwissProt_39:LAMP_HUMAN	+	66.00	135.11	2.21	338 ! Q13449 homo sapiens (human).
SwissProt_39:LAMP_RAT	+	66.00	135.11	2.21	338 ! Q62813 rattus norvegicus (rat).
SwissProt_39:LAMP_MOUSE	+	64.50	128.06	3.55	521 ! Q61790 mus musculus (mouse).
SwissProt_39:YJ9P_YEAST	+	64.50	121.37	3.75	1161 ! P47179 saccharomyces cerevisiae (yeast).
SwissProt_39:YC31_SCHPO	+	63.50	117.34	5.11	1429 ! Q14065 schizosaccharomyces (yeast).
SwissProt_39:SSF2_HUMAN	+	63.00	130.79	5.22	249 ! P28290 homo sapiens (human).
SwissProt_39:LAMP_CHICK	+	63.00	128.24	5.34	338 ! Q98919 gallus gallus (chicken).
SwissProt_39:SYA_ECOLI	+	63.00	120.29	5.72	876 ! P00957 escherichia coli.
SwissProt_39:GTFC_SPMU	+	63.00	116.52	5.90	1375 ! P13470 streptococcus mutans.
SwissProt_39:HFCL_MESAU	+	63.00	113.02	6.08	2090 ! P51611 mesocricetus auratus.
SwissProt_39:H11_MOUSE	+	62.00	119.85	6.93	212 ! P43275 mus musculus (mouse).
SwissProt_39:THII_ARCFU	+	62.00	125.11	7.21	374 ! Q29382 archaeoglobus fulgidus.
SwissProt_39:Y830_THEMEA	+	61.50	122.72	8.44	434 ! Q9wzt7 thermotoga maritima.
SwissProt_39:VGLH_HSV7J	+	61.50	118.85	8.73	690 ! P52353 human herpesvirus (ty).
SwissProt_39:YL51_CAEEL	+	61.00	109.95	10.80	1745 ! P34431 caenorhabditis elegans.
SwissProt_39:PRIS_ECOLI	-	60.00	117.31	13.34	550 ! P75825 escherichia coli.
SwissProt_39:TRBE_AGRTU	+	59.50	112.83	15.90	820 ! P54910 agrobacterium tumefaciens.
SwissProt_39:CRP2_RAT	+	59.00	123.14	16.69	208 ! P36201 rattus norvegicus (rat).
SwissProt_39:LOL2_HUMAN	+	59.00	112.16	18.33	774 ! Q9Y4K0 homo sapiens (human).
SwissProt_39:SYG1_MYCTU	+	58.50	115.20	20.49	469 ! P96862 mycobacterium tuberculosis.
SwissProt_39:Y075_MYCCE	+	58.50	108.68	21.66	1024 ! P47321 mycoplasma genitalium.
SwissProt_39:CRP2_HUMAN	+	58.00	120.85	22.39	208 ! P52943 homo sapiens (human).
SwissProt_39:MASY_RAPSA	+	58.00	112.49	24.05	566 ! Q43827 raphanus sativus (rad).
SwissProt_39:MASY_GOSHI	+	58.00	112.47	24.05	567 ! P17432 gossypium hirsutum (u).
SwissProt_39:SP13_VACCC	+	58.00	111.28	24.30	654 ! Q13674 schizosaccharomyces (yeast).
SwissProt_39:SP13_VACCC	+	57.50	114.92	27.01	369 ! P20532 vaccinia virus (strain).
SwissProt_39:SP13_VACCC	+	57.50	114.83	27.03	373 ! P33831 variola virus.
SwissProt_39:SP13_VACCC	+	57.50	112.25	27.64	508 ! Q28078 archaeoglobus fulgidus.
SwissProt_39:MXID_SHIFL	+	57.00	110.20	32.25	566 ! Q04641 shigella flexneri.
SwissProt_39:MXID_SHISO	+	57.00	110.20	32.25	566 ! Q52293 shigella sonnei.
SwissProt_39:UL32_HSV1J	+	57.00	109.77	32.37	596 ! P10216 herpes simplex virus.
SwissProt_39:RGSK_CHICK	-	56.50	117.02	34.90	218 ! Q9pwal gallus gallus (chicken).
SwissProt_39:HRCA_LACLA	-	56.50	113.14	36.07	347 ! Q9cgt0 lactococcus lactis (s).
SwissProt_39:SDC3_CHICK	+	56.50	111.85	36.47	405 ! P26261 gallus gallus (chicken).
SwissProt_39:HMT_DROME	+	56.50	111.69	36.52	413 ! P02835 drosophila melanogaster.
SwissProt_39:YFVB_METTF	+	56.50	110.95	36.75	451 ! P29578 methanobacterium thermautotrophicum.
SwissProt_39:SP21_YEAST	+	56.50	106.61	38.14	758 ! P35209 saccharomyces cerevisiae.
SwissProt_39:PEPX_LACDL	+	56.50	106.25	38.26	792 ! P40334 lactobacillus delbrueckii.
SwissProt_39:P3K1_DICDI	+	56.50	100.53	40.17	1570 ! P54673 dictyostellium discoideum.

alignment_scores:

Quality: 146.00

Length: 65

SwissProt_39:CRP1_HUMAN + 56.00 124.68 37.49 76 ! P50238 homo sapiens (human)
SwissProt_39:CRP1_MOUSE + 56.00 124.68 37.49 76 ! P04006 mus musculus (mouse)
SwissProt_39:YPSS_RHLP + 56.00 122.82 38.09 95 ! P10497 rhizobium leguminosarum
SwissProt_39:SECY-STACA + 56.00 110.21 42.42 430 ! Q05217 staphylococcus carnosus
SwissProt_39:GUX2_TIRE + 56.00 109.44 42.70 471 ! P07987 trichoderma reesei

seq_name: SwissProt_39:ECIC_BOVIN

seq_documentation_block:

ID ECIC_BOVIN STANDARD; PRT; 903 AA.
AC P54281;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EPIHELIAL CHLORIDE CHANNEL PROTEIN (CALCIUM-ACTIVATED CHLORIDE CHANNEL).
DE Bos taurus (Bovine).
OS Bos taurus (Bovine).
OC Pukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Trachea;
RX MEDLINE=96125078; PubMed=8537359;
RA Cunningham S.A., Awayda M.S., Buben J.K., Ismailov I.I., Arrate M.P., Berdiev B.K., Benos D.J., Fuller C.M.;
RT "Cloning of an epithelial chloride channel from bovine trachea.";
RL J. Biol. Chem. 270:31016-31026(1995).
CC !- FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME; MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND TRANSEPITHELIAL TRANSPORT.
CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC !- TISSUE SPECIFICITY: TRACHEA.
CC !- PTM: PHOSPHORYLATED BY CAM-KINASE II.
CC !- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC EMBL: U36445; AAC48511.1;
DR InterPro: IPR002035; VWFA.
DR SMART: SM00327; VWA; 1.
DR PROSITE: PS50234; VWFA; 1.
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel; Calcium channel; Phosphorylation; Glycoprotein.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 331 351 POTENTIAL.
FT TRANSMEM 617 637 POTENTIAL.
FT TRANSMEM 883 903 POTENTIAL.
FT DOMAIN 308 476 VWFA..
FT CARBOHYD 75 75 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 688 688 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 811 811 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 816 816 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 842 842 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 857 857 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 903 AA; 100305 MW; 9742AB1590908AEC CRC64;

FT CARBOHYD 244 244 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 521 AA; 56977 MW; 64C29CA0EB780E48 CRC64;

alignment_scores:
Quality: 64.50 Length: 46
Ratio: 2.224 Gaps: 1
Percent Similarity: 63.043 Percent Identity: 36.957

alignment_block:

US-09-049-696-11 x LAG3_MOUSE

Align seg 1/1 to: LAG3_MOUSE from: 1 to: 521

34 TACCTCCAAATCCAGGATGCTAGGTGGCAGTTGGAATACAGTCT 83

313 HisLeuGluAlaValcylLeuAlaGlnAlaGlyThrThrCysSer11 329

84 GCAAGCAAGCTCACAAACCTTCACCTGCTGCTACGTCCTCGGTGCTCCA 133

329 eHisLeuGlnGlyGlnGlnLeuAsnAlaThrVal..... 340

134 ATGCTACCTGCTGCTCCAAATACAGTACGTTCCAAACG 171

341ThrLeuAlaValIleThrValThrProLysSer 351

seq_name: SwissProt_39:YJ9P_YEAST

seq_documentation_block:

ID YJ9P_YEAST STANDARD; PRT; 1161 AA.

AC P47179;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE HYPOTHETICAL 118.4 KDA PROTEIN IN BAT2-DALS INTERGENIC REGION
DE PRECURSOR.

GN YJ151C OR J2223.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RA Scarcez T.;

RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR

CC (POTENTIAL)

CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE SRP1 / TIPI FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z49651; CAA89684.1; -

DR SGD; S0003912; YJ151C.

DR InterPro; IPR000992; SRP1_TIPI.

DR Pfam; PF00660; SRP1_TIPI; 1.

DR PROSITE; PS00724; SRP1_TIPI; 1.

KW Hypothetical protein; Glycoprotein; Signal; GPI-anchor.

FT SIGNAL 1 24 POTENTIAL.

FT CHAIN 25 1146 HYPOTHETICAL PROTEIN YJ151C.

FT PROPEP 1147 1161 REMOVED IN MATURE FORM (POTENTIAL).

FT LIPID 1146 1146 GPI-ANCHOR (POTENTIAL).

SQ SEQUENCE 1161 AA; 118358 MW; 7954C15D69F0CA58 CRC64;

alignment_scores:
Quality: 64.50 Length: 51
Ratio: 2.081 Gaps: 1
Percent Similarity: 60.784 Percent Identity: 37.255

alignment_block:

US-09-049-696-11 x YJ9P_YEAST

Align seg 1/1 to: YJ9P_YEAST from: 1 to: 1161

34 TACCTCCAAATCCAGGATGCTAGGTGGCAGTTGGAATACAGTCT 83

118 TyrThrAlaIleProThrSerThrSerThrThrThrLysSerSer.. 133

84 GCAAGCAAGCTCACAAACCTTCACCTGCTGCTACGTCCTCGGTGCTCCA 133

134ThrSerThrThrProThrThrThrIleThrSerThrThrSert 148

134 ATGCTACCTGCTGCTCCAAATACAGTACGTTCCAAACGACGACACC 183

148 hrThrSerThrThrProThrThrSerThrThrSerThrThrProThrThr 164

184 AGC 186

165 Ser 165

seq_name: SwissProt_39:YC31_SCHPO

seq_documentation_block:

ID YC31_SCHPO STANDARD; PRT; 1429 AA.

AC Q14065; P78884; Q9USG5;

DT 15-JUL-1998 (Rel. 36, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE HYPOTHETICAL 156.0 KDA PROTEIN C962.01 IN CHROMOSOME III.

GN SC962.01 OR SPC31B10.09

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OX NCBI_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RA Wedler H.; Dueterhoeft A.; McDougall R.C.; Rajandream M.A.;

RA Barrell B.G.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

RN [2] SEQUENCE OF 916-1429 FROM N.A.

RP STRAIN-972;

RA Harris D.; Wood V.; Rajandream M.A.; Barrell B.G.;

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases..

RN [3] SEQUENCE OF 1038-1429 FROM N.A.

RC STRAIN-PR745;

RA Yoshida S.; Kato K.; Okayama H.;

RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 2 C2 DOMAINS.

CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A

CC FRAMESHIFT IN POSITION 260.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AL121859; CAB58375.1; -

DR EMBL; AL031323; CAA20433.1; -

DR EMBL; D89235; BAA13896.1; ALT_FRAME.

DR FIR; JH0629; JH0629.
DR MIM; 118990;...
KW Sperm; Antigen.
SQ SEQUENCE 249 AA; 26821 MW; 15887378D3BAE509 CRC64;

'Align seg 1/1 to: SSF2_HUMAN from: 1 to: 249

3 CTTTGTAGTGGAACAAACACCACCAATGGCTACTCCAAATCCAGGCA 52
:::||||| ::||| ||| ::|||:
104 lleCysSerGlyProSerHisAlaAsn.ArgArgThrGlyValProSert 120

104 IleCysSerGlyProSerHisAlaAsn.ArgArgThrGlyValProSerT 120

.33 TTGCTAAGGTTGGCAGCTTGGAAATACAGTCTGCAAGCA..... 90
 : |||::||| || || || ||
 : |||::||| || || || ||

130 111A1ASEI VAI G I Y L Y S S E I D Y S I N I P R O T E U V A I A T A A R G L Y S V A I 136

91AGCTCACAACCTTGACCCCTGACTGTACAGTCCCGTGGCTCAA 134

137 PheArgAlaSerValAlaLeuThrProThrAlaProSerArgThrGlyse 153

```
.      .  
135 TGCTACCCTGCCTCCA 150  
.      .  
    .   .   .   .  
    :   :   :   |  
    :   :   :   |||
```

153 rValGlnThrProPro 158

```
seq_name: SwissProt_39:LAMP_CHICK
```

```
seq_documentation_block:
  id      LAMP CHICK  STANDARD:
```

AC Q98919;
DT 01-NOV-1997 (Rel. 3)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE LIMBIC SYSTEM-ASSOCIATED MEMBRANE PROTEIN PRECURSOR (E19S)
20-AUG-2001 (Rel. 40, Last annotation update)

CS Gallus gallus (Chicken).

CC	Corpus.
OX	NCBI_TaxID=9031;
BN	[1]

RP SEQUENCE FROM N.A.
RC TISSUE=Brain:

RX MEDLINE=97157768; PubMed=9004047;
FA Wilson D.J.A., Kim D.S., Clarke G.A., Marshall-Clarke S.,

RT "A family of glycoproteins (GP55), which inhibit neurite
RT are members of the Ig superfamily and are related to OBCAM
RT

RT neurotrophin, LAMP and CEPU-1.";
RL J. Cell Sci. 109:3129-3138(1996).

RN [2]
RP SEQUENCE FROM N.A.

RC TISSUE=Brain;
RX MEDLINE=97358596; PubMed=9215692;

RA Brummendorf T., Spaltmann F., Treubert U.;
 RF "Cloning and characterization of a neural cell recognition

RL Eur. J. Neurosci. 9:1105-1116(1997).

PROBABLY SERVES AS A RECOGNITION MOLECULE FOR THE FORMATION OF LAMINAR CONNECTIONS AND SELECTIVELY MEDIATES NEURONAL GROWTH AND AXON FUNCTION; MEDIATES SELECTIVE NEURONAL GROWTH AND AXON

CELL-TO-CELL CONNECTIONS (BY SIMILARITY).
CELL-TO-CELL LOCATION: ATTACHED TO THE MEMBRANE BY A GP
CELL-TO-CELL LOCATION: ATTACHED TO THE MEMBRANE BY A GP

- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
- SIMILARITY: BELONGS TO THE ICION FAMILY

[illegible][illegible]

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----

DR EMBL; Y08171; CAA69357.1; -
 DR EMBL; Z94720; CAB08115.1; -
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003598; Ig_C2.
 DR InterPro; IPR003600; Ig_Like.
 DR Pfam; PF00047; Ig; 3.
 DR SMART; SM00408; IGC2; 2.
 DR SMART; SM00410; IG_Like; 1.
 DR Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
 KW Signal. 1 28 POTENTIAL.
 FT CHAIN 29 338 LIMBIC SYSTEM-ASSOCIATED MEMBRANE
 FT PROPEP 29 338 REMOVED IN MATURE FORM (POTENTIAL).
 FT DOMAIN 46 118 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 146 204 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 232 297 IG-LIKE C2-TYPE DOMAIN 3.
 FT DISULFID 53 111 POTENTIAL.
 FT DISULFID 153 197 POTENTIAL.
 FT DISULFID 239 290 POTENTIAL.
 FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 287 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 338 AA; 37394 MW; 8FAA6A60AD98426B4 CRC64;

alignment_scores:

Quality: 63.00 Length: 48
 Ratio: 1.969 Gaps: 0
 Percent Similarity: 66.67 Percent Identity: 29.167

alignment_block:

US-09-049-696-11 x LAMP_CHICK ..
 Align seg 1/1 to: LAMP_CHICK from: 1 to: 338
 34 TACCTCCAAATCCAGCAATGCTAAGTTGGCAGCTTGGAAATACAGTCT 83
 181 TyrLeuGluIleLeuGlyIleThrArgGluGlnSerGlyLysTyrGluCy 197
 84 GCAAGCAAGCTCACAAACCTTGACCTGCTGCTGCTGCTGCTGCTGCTCA 133
 197 slysAlaAlaAsnGluValAlaSerAlaAspValLysGlnValArgValt 214
 134 ATGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 177
 214 hrValAsnTyrProThrIleThrGluSerLysSerAsnGlu 228

seq_name: SwissProt_39:SYA_ECOLI

seq_documentation_block:

ID SYA_ECOLI STANDARD; PRT; 876 AA.
 AC P00957; P78279;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ALANYL-TRNA SYNTHETASE (EC 6.1.1.7) (ALANINE--TRNA LIGASE) (ALARS).
 GN ALAS OR LOVB OR B2697.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.

OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 Mau B., Shao Y.;
 RA "The complete genome sequence of Escherichia coli K-12.";
 RT Science 277:1453-1474(1997).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97349980; PubMed=9205837;
 RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
 Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
 Oshima T., Oiyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
 Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 Yamagata S., Horiuchi T.;
 RA "Construction of a contiguous 874-kb sequence of the Escherichia coli
 RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113(1997).
 RN [3]
 RP SEQUENCE OF 1-166 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=81101084; PubMed=7005898;
 RA Herlihy W.C., Royal N.J., Biemann K., Putney S.D., Schimmel P.R.;
 RT "Mass spectra of partial protein hydrolysates as a multiple phase
 RT check for long polypeptides deduced from DNA sequences: NH2-terminal
 RT segment of alanine tRNA synthetase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 77:6531-6535(1980).
 RN [4]
 RP SEQUENCE OF 2-876 FROM N.A.
 RX MEDLINE=82017165; PubMed=7025207;
 RA Putney S.D., Royal N.J., de Vegvar H.N., Herlihy W.C., Biemann K.,
 Schimmel P.;
 RT "Primary structure of a large aminoacyl-tRNA synthetase.";
 RL Science 213:1497-1501(1981).
 RN [5]
 RP SEQUENCE OF 799-876 FROM N.A.
 RC STRAIN=KP4714;
 RX MEDLINE=96177756; PubMed=8604133;
 RA Murayama N., Shimizu H., Takiguchi S., Baba Y., Amino H., Horiuchi T.,
 Sekimizu K., Miki T.;
 RA "Evidence for involvement of Escherichia coli genes pmBA, csrA and a
 RT previously unrecognized gene tldD, in the control of DNA gyrase by
 RT leuD (ccdB) of sex factor F.";
 RL J. Mol. Biol. 256:483-502(1996).
 RN [6]
 RP SEQUENCE OF 866-876 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=93328679; PubMed=8393005;
 RA Romeo T., Gong M., Liu M.-Y., Brun-Zinkernagel A.-M.;
 RT "Identification and molecular characterization of csrA, a pleiotropic
 RT gene from Escherichia coli that affects glycogen biosynthesis,
 RT gluconeogenesis, cell size, and surface properties.";
 RL J. Bacteriol. 175:4744-4755(1993).
 RN [7]
 RP SEQUENCE OF 1-13 FROM N.A.
 RX MEDLINE=81093926; PubMed=6256345;
 RA Putney S.D., Melendez D.L., Schimmel P.R.;
 RT "Cloning, partial sequencing, and in vitro transcription of the gene
 RT for alanine tRNA synthetase.";
 RL J. Biol. Chem. 256:205-211(1981).
 RN [8]
 RP ZINC-BINDING.
 RX MEDLINE=91299753; PubMed=1712632;
 RA Miller W.T., Hill K.A.W., Schimmel P.;
 RT "Evidence for a 'cysteine-histidine box' metal-binding site in an
 RT Escherichia coli aminoacyl-tRNA synthetase.";
 RL Biochemistry 30:6970-6976(1991).


```
seq_documentation_block:
AC THII_ARCFU STANDARD; PRT; 374 AA.
O29382;
DT 15-DEC-1998 (Rel. 37, Created)
LT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE THIAMINE BIOSYNTHESIS PROTEIN THII.
GN THII OR AF0879.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OX Archaeoglobus.
NCBI_TaxID=2234;
[!]
RN SEQUENCE FROM N.A.
RP STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpidides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
PA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
PA Patterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RA "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- FUNCTION: REQUIRED FOR THE SYNTHESIS OF THE THIAZOLE MOIETY (BY
SIMILARITY).
CC -!- PATHWAY: THIAMINE BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE THII FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
-----
CC EMBL; AE001043; AAB90361.1; -.
DR TIGR; AF0879; .
DR InterPro; IPR003720; ThII.
DR Pfam; PF02568; ThII; 1.
KW Thiamine biosynthesis; Complete proteome.
SQ SEQUENCE 374 AA; 42272 MW; AA3C0833ABI7FD98 CRC64;
-----
alignment_scores:
.. Quality: 62.00 Length: 55
Ratio: 1.938 Gaps: 3
Percent Similarity: 58.182 Percent Identity: 38.182
-----
alignment_block:
US-09-049-696-11/rev x THII_ARCFU ..
Align seg 1/1 to: THII_ARCFU from: 1 to: 374
-----
191 AATTGCGGTGGTGCTCTTCCTTTTGAAGCACTGTAATTTGGAGGCAG 142
||| ||||| :||||| :||| |||
: 114 AsnArgLysValGlyAlaPhellevalGlulysThr.....GlyLysLy 128
141 GTTAGCATPTGGACGCACGGGAGTCGACAGTCAGGTCGAAGGTCTTGTGAGC 92
:||| ||| | ||||| :||| :||| :||| :|||
128 sValAspLeuthrAsnProAspValThrValTrpIleGlulcycsGIuL 145
CC TTGCTWTGACACTGTATTTCAAAGTGCCCAACTTAGCAATGCCCTGGGATT 42
::: ::| ||| |||||
145 ySGluAlaTyrrValtyr_SerLys.....ArqTvriGluGIuVII 157
```

```
41 TCGAGGTAGGCCA 29
   :|||||
157 eglyGlyLeuPro 161

seq_name: SwissProt_39:Y830_THEME
seq_documentation_block:
ID Y830_THEME STANDARD; PRT; 434 AA.
AC Q9WZT7;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN TM0830.
GN TM0830.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=9287316; PubMed=10360571;
RA Nelson K.E., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA Haft D.H., Hickey E.K., Peterson J.D., Linher K.D., Garrett M.M.,
RA McDonald L., Utterback T.R., Malek J.A., Phillips C.A., Richardson D.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 395:323-329(1999).
CC -!- SIMILARITY: BELONGS TO THE UPF0004 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF001750; AAD35912.1; -
DR TIGR; TM0830; -
DR InterPro; IPR001861; UPF0004.
DR Pfam; PF00919; UPF0004; 1.
DR PROSITE; PS01278; UPF0004; 1.
KW Hypothetical protein; Complete proteome.
SQ
SEQUENCE 434 AA; 49324 MW; C33186CD10545B20 CRC64;

alignment_scores:
Quality: 61.50 Length: 65
Ratio: 2.050 Gaps: 2
Percent Similarity: 46.154 Percent Identity: 26.154

alignment_block:
US-09-049-696-11 x Y830_THEME ..
Align seg 1/1 to: Y830_THEME from: 1 to: 434

42 AATCCAGCATTCGTAAGGTGGCACTTGGAAATACAGTCTGCAAGC.. 89
|||||
242 AsnProArgLeuCysProHisLeuHisIleSerValGlnSerGlySerAs 258
89 .....
258 AspValLeuLysArgMetGlyArgLysTyrLysIleSerAspPheMetA 275
90 .....
275 rgValValAspLysLeuArgSerIleAspProAspPheSerIleThr 291
123 CGGTGCGTCCAAATGCTACCTCGCTCCCAATTAACAGTGCATTCCAA 167
```

119 CGTCCGCTGCCATCAAGCTACCTGGCTCCAATTACAGTGACT...TTC 165
||| ::: ||::: ||::: ||| :::
557 leSerThrIleIleIleThrValProLeuAsnSerThrCysThr 573

166 AAAACGAAC 174
||||
574 ProThrAsn 576

This Page Blank (uspio)

OM of: US-09-049-696-11 to: PIR_68:* out_format : pfs
 Date: Mar 30, 2002 2:27 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+g2p.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09049696/runat_28032002_145238_2031/app_query.fasta_1.12579
 -DB=PIR_68 -OFMT=fastan -SUFFX=pr -GAPO=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000 -OGAPO=4.500
 -OGAPEXT=0.050 -XGAPO=10.000 -XGAPEXT=0.500 -FGAPO=6.000
 -FGAPEXT=7.000 -YGAPO=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=dlosum62 -TRANS=human40.cdi
 -LIST=45 -DALIGN=200 -THR_SCORE=pc -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=200000000 -USER=US09049696 @CGN1_1_367 -NCPU=6 -ICPU=3
 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-049-696-11
 Query length: 191
 Database: PIR_68:*
 Database sequences: 219241
 Database length: 76174552
 Search time (sec): 470.790000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
pir2:7020168	+	246.00	3.1e-22	913	! gob-5 protein - mouse
pir2:702205	+	153.00	3.1e-10	905	! Lu-ECAM-1 protein - mouse
pir2:681716	+	70.00	139.63	397	! hypothetical protein TC0320 [im
pir2:706000	+	66.50	131.42	420	! aspartic proteinase homolog p17
pir2:704776	+	66.00	131.91	338	! limbic-system-associated membra
pir2:857180	+	64.50	119.46	1161	! probable membrane protein YUK1
pir2:743139	+	63.50	127.45	291	! hypothetical protein - fission
pir2:741848	+	63.50	123.25	514	! C2 domain family protein - fis
pir2:741699	+	63.50	115.70	1429	! C2-domain family protein - fis
pir2:706029	+	63.00	127.49	249	! cleavage signal-1 protein - hu
pir2:705519	+	63.00	125.23	338	! 50K glycoprotein precursor - ch
pir2:86798	+	63.00	124.50	373	! prophage p13 protein 11 [import
pir2:703146	+	63.00	120.04	683	! probable glycoprotein A8 - alce
pir2:70345	+	63.00	118.20	876	! alanine--trna ligase (EC 6.1.1.1
pir2:738271	+	63.00	114.88	1375	! dextranucrase (EC 2.4.1.5) pr
pir2:703217	+	62.50	122.45	424	! conserved hypothetical protein
pir2:734369	+	62.50	115.57	1076	! major surface glycoprotein 5 -
pir2:843949	+	62.50	111.87	1777	! hypothetical protein T19D12.1
pir2:72479	+	62.00	126.41	213	! histone H1 - mouse
pir2:72479	+	62.00	122.68	353	! hypothetical protein APE2475 -
pir2:72479	+	62.00	122.26	374	! conserved hypothetical protein
pir2:72479	+	62.00	109.33	374	! proline-rich protein Rad9 - in
pir2:72479	+	61.50	121.70	347	! adhesion p1 precursor homolog RQ
pir2:72479	+	61.50	121.03	380	! Ig heavy chain precursor - lit
pir2:72479	+	61.50	120.05	434	! conserved hypothetical protein
pir2:72479	+	61.50	118.61	527	! adhesion p1 precursor homolog PC
pir2:72479	+	61.50	116.62	690	! glycoprotein H - human herpesv
pir2:72479	+	61.50	111.16	1446	! carboxypeptidase D (EC 3.4.17.
pir2:72479	+	61.50	101.73	5188	! probable RFX family exoprotein
pir2:72479	+	61.00	112.72	27.26	! hypothetical protein T07C12.8
pir2:72479	+	61.00	108.66	26.47	! F44E2.1 protein - Caenorhabdit
pir2:72479	+	61.00	106.72	26.10	! gag, pol and env protein precu
pir2:72479	+	60.50	113.39	31.85	! hypothetical protein SPAC21B10.
pir2:72479	+	60.00	126.01	40.58	! hypothetical protein Y65A5A.1 -
pir2:72479	+	60.00	117.55	38.17	! heme d1 biosynthesis protein Ni
pir2:72479	+	60.00	114.96	37.46	! hybrid cluster [4fe-2s-3o] prot
pir2:72479	+	60.00	114.93	37.45	! probable prismase ybJW [Importe
pir2:72479	+	59.50	115.10	43.60	! hypothetical protein T23F4.4 -
pir2:72479	+	59.50	112.80	42.91	! hypothetical protein T1F3.6 [I
pir2:72479	+	59.50	112.88	42.91	! Ig mu chain precursor, membra
pir2:72479	+	59.00	119.91	52.49	! cysteine-rich protein CRP2 - ra

pir2:T18198 + 59.00 114.05 50.32 460 ! hypothetical protein 1 - sil
 pir2:T23175 - 59.00 111.50 49.40 650 ! hypothetical protein K01C8.3
 pir2:E85917 - 59.00 109.28 48.61 878 ! alanyl-trna synthetase [impo
 pir2:S64916 + 59.00 106.85 47.76 1220 ! probable membrane protein y

seq_name: pir2:JG0168

seq_documentation_block:

gob-5 protein - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
 C:Accession: JG0168

R.Komiya, T.; Tanigawa, Y.; Hirohashi, S.

Biochem. Biophys. Res. Commun. 255, 347-351, 1999

A:Title: Cloning and identification of the gene gob-5, which is expressed in intestin

A:Reference number: JG0168; MUID:99160866

A:Accession: JG0168

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-913 <KOM>

A:Cross-references: DDBJ:AB016592

alignment_scores:

Quality: 246.00 Length: 63
 Ratio: 4.241 Gaps: 0
 Percent Similarity: 92.063 Percent Identity: 74.603

alignment_block:

US-09-049-696-11 x JG0168

Align seg 1/1 to: JG0168 from: 1 to: 913

1 GGCTTTTGTAGTGGACAAAACACCAAAATGGCTTACCTCCAAATCCAGG 50
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 542 GlypheileLeuaspThrThrThrLysValAlaThrLeuGlnValProG 558

51 CATTGTAAAGTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCAGAAA 100
 | |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 558 yThrAlaLysValGlyPheTrpLysTrpSerIleGlnAlaSerSergInt 575

101 CCTTGACCCGTCGTCCACCTCCGTCGTCCTCAATGCTACCTGCCCTCCA 150
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 575 hrLeuThrLeuThrValThrSerArgAlaAlaThrLeuProPro 591

151 ATTACAGTACTTCCAAAACGACCAAGACACCAAGCAAGCAAGCAAGCA 189

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

592 IleThrValThrProValValAsnLysAsnThrGlyLys 604

seq_name: pir2:T02205

seq_documentation_block:

Lu-ECAM-1 protein - bovine

C:Species: Bos primigenius taurus (cattle)

C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 01-Dec-2000

C:Accession: T02205; T02152; T02171

R.Elbert, R.C.; Widom, J.; Gruber, A.D.; Abdel-Ghany, M.; Levine, R.; Goodwin, A.; Pau

submitted to the EMBL Data Library, April 1997

A:Description: Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial

A:Reference number: Z14590

A:Accession: T02205

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-905 <ELB>

A:Cross-references: EMBL:AF001261; MID:g2623762; PIDN:AAB86529.1; PID:g2623763

A:Experimental source: lung

A:Accession: T02152

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-793; 'SGSFP', 'M', '862', 'RF', '865-866', 'Q', '868', 'AKVLEIQ', '876', 'QHO', '880', 'FQ' <

A:Cross-references: EMBL:AF001263; MID:g2623766; PIDN:AAB86531.1; PID:g2623767

A:Experimental source: lung

A:Accession: T02171

|||||:||||| |||||||:||||| |||||||:||||| |||||||:||||| |||||||:|||||
 181 TyrLeuGluIleLeuGlyIleThrArgGluGlnSerGlyLysTyrGluCy 197
 84 GCAAGCAAGCTCAACAACCTTGACCTGACTGTACACGTCCCGTGGCGTCCA 133
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 197 sLysAlaAlaAsnGluValAlaSerAlaAspValLysGlnValArgValt 214
 134 ATGCTACCTCGCTCCCAATACAGTGACTTCCAAACGACACAG 177
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 214 hrValAsnTyrProThrIleThrGluSerLysSerAsnGlu 228

seq_name: pir2:A86798

seq_documentation_block:

prophage pi3 protein 11 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis

C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001

C:Accession: A86798

R:Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich

Genome Res. in press, 2001

A:Title: The complete genome sequence of the lactic acid bacterium.

A:Reference number: A86625

A:Accession: A86798

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-373 <STO>

A:Cross-references: GB:AE005176; NID:g12724371; PIDN:AAK05483.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: pi311

alignment_scores:

Quality: 63.00 Length: 73
 Ratio: 1.750 Gaps: 2
 Percent Similarity: 49.315 Percent Identity: 26.027

alignment_block:

US-09-049-696-11 x A86798 ..

Align seg 1/1 to: A86798 from: 1 to: 373

31 GCCTACCTCCCAATCCAGGCATT.....GCTAAGT 62
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||
 22 AlaphellethrAlaSerGlyValThrThrGlnThrThrAlaThrAl 38
 63 TGGCAGCTTGGAAATACAGTCTGCAAGCAAGCTCACAACTTGACCTGA 112
 : |||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||
 38 aGlnThrTrpSerProSerTyrAlaThrThrProGlnValLeuThrLeuA 55
 113 CTGTCAGCTCCGCTGCTCCCAATGCTACCTG..... 144
 :|||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||
 55 snLeuThrLysAlaGlySerThrThrSerValIleGlyIleSerGly 71
 145CCTCCAATTACAGTGACTTC 164
 72 AsnIleThrTrpThrArgAlaAspGlyThrThrThrThrIleThrse 88

165 CAAACGACGACACACC 183

I ||||| |||||

88 rThrThrAsnThrAspThr 94

seq_name: pir2:T03146

seq_documentation_block:

probable glycoprotein A8 - alcelaphine herpesvirus 1

C:Species: alcelaphine herpesvirus 1

C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999

C:Accession: T03146

R:Ensser, A.; Pflanz, R.; Fleckenstein, B.

J. Virol. 71, 6517-6525, 1997

A:Title: Primary structure of the alcelaphine herpesvirus 1 genome.

A:Reference number: Z14840; MUID:97404659

A:Accession: T03146
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-683 <ENS>
 A:Cross-references: EMBL:AF005370; NID:g2337967; PIDN:AAK58098.1; PID:g2338014

alignment_scores:

Quality: 63.00 Length: 63
 Ratio: 1.750 Gaps: 1
 Percent Similarity: 57.143 Percent Identity: 30.159

alignment_block:

US-09-049-696-11 x T03146 ..

Align seg 1/1 to: T03146 from: 1 to: 683

13 GACAAAACACCAAAATGGCTACCTCCAATCCAGGC..... 51

|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

484 AsHisLysThrValSerProSerLeuValThrProGlyArgThrSerTh 500

52 .ATTGCTAGGTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAAA 100

|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

500 rLeuProIleValSerMetThrHisPheSerArgGluGlySerProL 517

101 CCTTGACCTGACTGTCACGTCGCGTCCCAATGCTACCTGCTGCCA 150

|||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||

517 ysProGlnThrThrAlaAlaLysThrSerSerGluAlaSerLeuPro 533

151 ATTACAGTGTCTCCAAACGACCAAGCACCCAGCAAA 189

|||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

534 LeuLeuThrThrThrProThrProThrProThrAsnThrGluLys 546

seq_name: pir1:SYECAT

seq_documentation_block:

alanine-tRNA ligase (EC 6.1.1.7) [validated] - Escherichia coli

N:Alternate names: alanyl-tRNA synthetase

C:Species: Escherichia coli

C>Date: 31-Mar-1981 #sequence_revision 31-Oct-1997 #text_change 02-Jun-2000

A:Accession: E65049; A01185; A94258; A40608; I40988; A54217

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: E65049

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-876 <BLAT>

A:Cross-references: GB:AE000353; GB:U00096; NID:g1789037; PIDN:AAK75739.1; PID:g17890

A:Experimental source: strain K-12, substrain MG1655

R:Herlihy, W.C.; Royal, N.J.; Biemann, K.; Putney, S.D.; Schimmel, P.R.

Proc. Natl. Acad. Sci. U.S.A. 77, 6531-6535, 1980

A:Title: Mass spectra of partial protein hydrolysates as a multiple phase check for 1

A:Reference number: A93864; MUID:81101084

A:Accession: A01185

A:Molecule type: DNA

A:Residues: 1-28,'RY',31,'IT',34-157,'N',159-166 <HER>

R:Putney, S.D.; Royal, N.J.; de Vegvar, H.N.; Herlihy, W.C.; Biemann, K.; Schimmel, P

Science 213, 1497-1501, 1981

A:Title: Primary structure of a large aminoacyl-tRNA synthetase.

A:Reference number: A94258; MUID:82017165

A:Accession: A94258

A:Molecule type: DNA

A:Residues: 2-28,'RY',31,'IT',34-157,'N',159-167,'G',169-171,'R',173-174,'G',176-179,

A:Cross-references: GB:J01581; GB:J01582; NID:9414793; PIDN:AAA03208.1; PID:g145220

R:Romeo, T.; Gong, M.; Liu, M.Y.; Brun-Zinkernagel, A.M.

J. Bacteriol. 175, 4744-4755, 1993

A:Title: Identification and molecular characterization of csra, a pleiotropic gene fr

A:Reference number: A40608; MUID:93328679

A:Accession: A40608

A:Molecule type: DNA

84 GCAGCAAGCTCACAAACCTTGACCTGACTGTCACGTCCGTCGCTCCA 133
:|||||: : : : :
169 slysaAlaAsnGluValSerSerAlaAspValLysValt 186
:|||||: : : : :
134 ATGCTACCTGCTCCCAATTACAGTCACTTCCAAACGAACAAGGACACC 183
:|||||: : : : :
186 hrValAsnTyProThrIleThrGluSerLysSerAsnGluAlaThr 202
:|||||: : : : :
184 AGC 186
: : :
203 Thr 203

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-414-657D-49

seq_documentation_block:
; Sequence 49, Application US/08414657D
; Patent No. 5861283
; GENERAL INFORMATION:
; APPLICANT: Levitt, Pat
; APPLICANT: Pimenta, Aurea
; APPLICANT: Fischer, Itzhak
; APPLICANT: Zhukareva, Victoria
; TITLE OF INVENTION: Limbic System-Associated Membrane
; TITLE OF INVENTION: Protein and DNA
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,657D
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-414-657D-49

alignment_scores:
Quality: 66.00 Length: 51
Ratio: 1.941 Gaps: 0
Percent Similarity: 66.667 Percent Identity: 29.412

alignment_block:
US-09-049-696-11 x US-08-414-657D-49
Align seg 1/1 to: US-08-414-657D-49 from: 1 to: 287

34 TACCTCCAATCCAGCATTGCTAGGTGGCACTGGAAATACAGTCT 83
|||||: : : : :
: : : : :
: : : : :

153 TyrLeuGluIleLeuGlyIleThrArgGluGlnSerGlyLysTyrGluCy 169
:|||||: : : : :
84 GCAAGCAAGCTCACAAACCTTGACCTGACTGTCACGTCCGTCGCTCCA 133
:|||||: : : : :
169 slysaAlaAsnGluValSerSerAlaAspValLysValt 186
:|||||: : : : :
134 ATGCTACCTGCTCCCAATTACAGTCACTTCCAAACGAACAAGGACACC 183
:|||||: : : : :
186 hrValAsnTyProThrIleThrGluSerLysSerAsnGluAlaThr 202
:|||||: : : : :
184 AGC 186
: : :
203 Thr 203

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-414-657D-44

seq_documentation_block:
; Sequence 44, Application US/08414657D
; Patent No. 5861283
; GENERAL INFORMATION:
; APPLICANT: Levitt, Pat
; APPLICANT: Pimenta, Aurea
; APPLICANT: Fischer, Itzhak
; APPLICANT: Zhukareva, Victoria
; TITLE OF INVENTION: Limbic System-Associated Membrane
; TITLE OF INVENTION: Protein and DNA
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,657D
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 304 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-414-657D-44

alignment_scores:
Quality: 66.00 Length: 51
Ratio: 1.941 Gaps: 0
Percent Similarity: 66.667 Percent Identity: 29.412

alignment_block:
US-09-049-696-11 x US-08-414-657D-44
Align seg 1/1 to: US-08-414-657D-44 from: 1 to: 304

34 TACCTCCAAATCCCAGGCAATTGGTCAAGTTGGGCACCTTGGAATACAGTCT 83
|||||:||||| |||||:||||| :||| :|||||:
153 TyrLeuGluIleLeuGlyIleThrArgcLgLnSerGlyLysrGluCy 169
84 GCAAGCAGCTCACAAAACCTTGACCCCTCACTGTCACTGCCCGTCGCTCCA 133
:|||:|||||:|||||:||||| :||| :||| :|||:
169 slysAlaAAsnGluValSerSerAlaAspValLysGlnValLysvalr 186
134 ATGCTACCCGTGCTCCCAATTACACTGACTTCCAAACGACAACAGGACACC 183
:|||||:||||| :||| :||| :||| :|||:
186 hrValasnTyrrProPthrIlehrGluSerLysSerAsnGluAlaThr 202

184 AGC 186
:||
203 Thr 203

seq_name: /cqn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-414-657D-46

seq_documentation_block:

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

```

alignment_scores:
  Quality: 66.00      Length: 51
  Ratio: 1.941      Gaps: 0
  percent Similarity: 66.667      Percent Identity: 29.412

alignment_block:
  US-09-049-696-11 x US-08-414-657D-46
  ..

```

```

Align seq 1/1 to: US-08-414-657D-46 from: 1 to: 308
|
| 34 TACTCTCAAATCCCGAGCATTCGTAAAGTTGGCAGCTTGGAAATACAGTCT 83
|      |||||:|||| |||||:|||||: |||||: |||||: |||||: |||||:
| 174 TyrLeuGluIleLeuGlyIleThrArgGluGlnSerGlyLysTyrGluCy 190
|
| 84 GCAAGCAAGCTCACAAACCTTGACCTGACTGTCAGTCCCGTGGGTCCA 133
|      |||||:|||||: |||||: |||||: |||||: |||||: |||||:
| 190 sLySAIaIAAsnGluValSerAlaAspValLysGlnValLysValT 207
|
| 134 ATGCTACCTCGCTCCCAATTCAGCTGACTTCCAAAAACGAACGACACC 183
|      |||||: |||||: |||||: |||||: |||||: |||||: |||||:
| 207 hrValAsnTyrProProThrIleThrGluSerLysSerAsnGluAlaThr 223
|
| 184 AGC 186 f.
|      :||:
| 224 Thr 224

```

seq name: ./cqn2 6/ptodata/2/iaa/5B COMB.pep:US-08-414-657D-45

```

seq_documentation_block:
? Sequence 45, Application US/08414657D
? Patent No. 5861283
? GENERAL INFORMATION:
? APPLICANT: Levitt, Pat
? APPLICANT: Pimenta, Aurea
? APPLICANT: Fischer, Itzhak
? APPLICANT: Zhukareva, Victoria
? TITLE OF INVENTION: Limbic System-Associated Membrane
? TITLE OF INVENTION: Protein and DNA
? NUMBER OF SEQUENCES: 60
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Dechert Price & Rhoads
? STREET: 997 Lenox Drive, Building 3, Suite 210
? CITY: Lawrenceville
? STATE: NJ
? COUNTRY: USA
? ZIP: 08543
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FASTSEQ for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/414,657D
? FILING DATE: 31-MAR-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Bloom, Allen
? REGISTRATION NUMBER: 29,135
? REFERENCE/DOCKET NUMBER: 317743-102
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 609-520-3214
? TELEFAX: 609-520-3259
? TELEX:
? INFORMATION FOR SEQ ID NO: 45:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 310 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? US-08-414-657D-45

```

alignment_scores:	
Quality:	66.00
Ratio:	1.941
Percent Similarity:	66.667
alignment_block:	
	Length: 51
	Gaps: 0
	Percent Identity: 29.412

US-09-049-696-11 x US-08-414-657D-45

Align seg 1/1 to: US-08-414-657D-45 from: 1 to: 310

```

34 TACCTCCAAATCCAGCATTGCTAAGTTGGCACTTGGAAATACAGTCT 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 TyrLeuGluIleLeuGlyIleThrArgGluGlnSerGlyLysTyrGluCy 169
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 GCAAGCAAGCTCACAAACCTTGACCTGACGTGCACGTCCCGTCCGTC 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 sLysAlaAlaAsnGluValSerAlaAspValLysGlnValLysValt 186
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 ATGCTACCTCGCTCCAAATTTACAGTGCACCTCCAAACGACAGGACAC 183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 hrValAsnTyrProThrIleThrGluSerLysSerAsnGluAlaThr 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 AGC 186
:::
203 Thr 203

```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-414-657D-47

```

seq_documentation_block:
; Sequence 47, Application US/08414657D
; Patent No. 5861283
; GENERAL INFORMATION:
; APPLICANT: Levitt, Pat
; APPLICANT: Pimenta, Aurea
; APPLICANT: Fischer, Itzhak
; APPLICANT: Zhukareva, Victoria
; TITLE OF INVENTION: Limbic System-Associated Membrane
; TITLE OF INVENTION: Protein and DNA
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,657D
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 315 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-414-657D-47

```

alignment_scores:
 Quality: 66.00 Length: 51
 Ratio: 1.941 Gaps: 0
 Percent Similarity: 66.667 Percent Identity: 29.412

alignment_block:

US-09-049-696-11 x US-08-414-657D-47

Align seg 1/1 to: US-08-414-657D-47 from: 1 to: 315

```

34 TACCTCCAAATCCAGCATTGCTAAGTTGGCACTTGGAAATACAGTCT 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 TyrLeuGluIleLeuGlyIleThrArgGluGlnSerGlyLysTyrGluCy 197
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 GCAAGCAAGCTCACAAACCTTGACCTGACGTGCACGTCCCGTCCGTC 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
197 sLysAlaAlaAsnGluValSerAlaAspValLysGlnValLysValt 214
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 ATGCTACCTCGCTCCAAATTTACAGTGCACCTCCAAACGACAGGACAC 183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 hrValAsnTyrProThrIleThrGluSerLysSerAsnGluAlaThr 230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 AGC 186
:::
231 Thr 231

```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-414-657D-2

```

seq_documentation_block:
; Sequence 2, Application US/08414657D
; Patent No. 5861283
; GENERAL INFORMATION:
; APPLICANT: Levitt, Pat
; APPLICANT: Pimenta, Aurea
; APPLICANT: Fischer, Itzhak
; APPLICANT: Zhukareva, Victoria
; TITLE OF INVENTION: Limbic System-Associated Membrane
; TITLE OF INVENTION: Protein and DNA
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,657D
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-414-657D-2

```

```
alignment_scores:
  Quality: 66.00      Length: 51
  Ratio: 1.941        Gaps: 0
Percent Similarity: 66.667 Percent Identity: 29.412

alignment_block:
US-09-049-696-11 x US-08-414-657D-2 ..

Align seg 1/1 to: US-08-414-657D-2 from: 1 to: 325

34 TACCTCCAAATCCAGGAGGATTCGTAAGCTTGGCACTTGGAAATACAGTCT 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 TyrLeuGluIleLeuGlyIleThrArgGluGlnSerGlyLysTyrGluCy 190
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 GCAAGCAAGCTCACAAACCTTGACCTGACTGTACGTCACGTCGCGTGCGTCCA 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
190 sLysAlaAlaAsnGluValSerSerAlaAspValLysGlnVallysValT 207
134 ATGCTACCTGCCTCCATTACAGTACTGCTCCAAACGAAACGACGACACC 183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 hrValAsnTyrProProThrIleThrGluSerLysSerAsnGluAlaThr 223
184 AGC 186
...
224 Thr 224

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.US-08-414-657D-41
```

```
seq_documentation_block:
; Sequence 41, Application US/08414657D
; Patent No. 5861283
; GENERAL INFORMATION:
; APPLICANT: Levitt, Pat
; APPLICANT: Pimenta, Aurea
; APPLICANT: Fischer, Itzhak
; APPLICANT: Zhukareva, Victoria
; TITLE OF INVENTION: Limbic System-Associated Membrane
; TITLE OF INVENTION: Protein and DNA
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,657D
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-414-657D-41
```

```
alignment_scores:
  Quality: 66.00      Length: 51
  Ratio: 1.941        Gaps: 0
Percent Similarity: 66.667 Percent Identity: 29.412

alignment_block:
US-09-049-696-11 x US-08-414-657D-41 ..

Align seg 1/1 to: US-08-414-657D-41 from: 1 to: 325

34 TACCTCCAAATCCAGGAGGATTCGTAAGCTTGGCACTTGGAAATACAGTCT 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 TyrLeuGluIleLeuGlyIleThrArgGluGlnSerGlyLysTyrGluCy 190
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 GCAAGCAAGCTCACAAACCTTGACCTGACTGTACGTCACGTCGCGTGCGTCCA 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
190 sLysAlaAlaAsnGluValSerSerAlaAspValLysGlnVallysValT 207
134 ATGCTACCTGCCTCCATTACAGTACTGCTCCAAACGAAACGACGACACC 183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 hrValAsnTyrProProThrIleThrGluSerLysSerAsnGluAlaThr 223
184 AGC 186
...
224 Thr 224

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.US-08-414-657D-42
```

```
seq_documentation_block:
; Sequence 42, Application US/08414657D
; Patent No. 5861283
; GENERAL INFORMATION:
; APPLICANT: Levitt, Pat
; APPLICANT: Pimenta, Aurea
; APPLICANT: Fischer, Itzhak
; APPLICANT: Zhukareva, Victoria
; TITLE OF INVENTION: Limbic System-Associated Membrane
; TITLE OF INVENTION: Protein and DNA
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,657D
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
US-08-414-657D-42
```



```
;
; TOPOLOGY: linear
; US-08-414-657D-42

alignment_scores:
  Quality: 66.00      Length: 51
  Ratio: 1.941       Gaps: 0
  Percent Similarity: 66.667   Percent Identity: 29.412

alignment_block:
  US-09-049-696-11 x US-08-414-657D-42

  Align seg 1/1 to: US-08-414-657D-42 from: 1 to: 338

34 TACCTCCAAATCCAGCATTCCTAAGGTGGCAGTTGGAATACAGTCT 83
|||||:||||| |||||:|||||: |||||:
181 TyrLeuGluIleLeuGlyIleThrArgGluGlnSerGlyLysTyrGluCy 197
|||||:||||| |||||:|||||: |||||:
84 GCAAGCAAGCTCACAAACCTTGACCTGCTCAGTCCCGTCCGTC 133
|||||:||||| |||||:|||||: |||||:
197 sLysAlaAlaAsnGluValSerSerAlaAspValLysGlnValLysValT 214
|||||:||||| |||||:|||||: |||||:
134 ATGCTACCTCGCTCCAAATACAGTGTCCAAACGAAGACAGCAC 183
|||||:||||| |||||:|||||: |||||:
214 hrValAsnTyrProThrIleThrGluSerLysSerAsnGluAlaThr 230
184 AGC 186
231 Thr 231

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.us-08-414-657D-43

seq_documentation_block:
; Sequence 43, Application US/08414657D
; Patent No. 5861283
; GENERAL INFORMATION:
; APPLICANT: Levitt, Pat
; APPLICANT: Pimenta, Aurea
; APPLICANT: Fischer, Itzhak
; APPLICANT: Zhukareva, Victoria
; TITLE OF INVENTION: Limbic System-Associated Membrane
; TITLE OF INVENTION: Protein and DNA
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,657D
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
```

```
;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-414-657D-43

alignment_scores:
  Quality: 66.00      Length: 51
  Ratio: 1.941       Gaps: 0
  Percent Similarity: 66.667   Percent Identity: 29.412

alignment_block:
  US-09-049-696-11 x US-08-414-657D-43

  Align seg 1/1 to: US-08-414-657D-43 from: 1 to: 338

34 TACCTCCAAATCCAGCATTCCTAAGGTGGCAGTTGGAATACAGTCT 83
|||||:||||| |||||:|||||: |||||:
181 TyrLeuGluIleLeuGlyIleThrArgGluGlnSerGlyLysTyrGluCy 197
|||||:||||| |||||:|||||: |||||:
84 GCAAGCAAGCTCACAAACCTTGACCTGCTCAGTCCCGTCCGTC 133
|||||:||||| |||||:|||||: |||||:
197 sLysAlaAlaAsnGluValSerSerAlaAspValLysGlnValLysValT 214
|||||:||||| |||||:|||||: |||||:
134 ATGCTACCTCGCTCCAAATACAGTGTCCAAACGAAGACAGCAC 183
|||||:||||| |||||:|||||: |||||:
214 hrValAsnTyrProThrIleThrGluSerLysSerAsnGluAlaThr 230
184 AGC 186
231 Thr 231

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.us-08-414-657D-60

seq_documentation_block:
; Sequence 60, Application US/08414657D
; Patent No. 5861283
; GENERAL INFORMATION:
; APPLICANT: Levitt, Pat
; APPLICANT: Pimenta, Aurea
; APPLICANT: Fischer, Itzhak
; APPLICANT: Zhukareva, Victoria
; TITLE OF INVENTION: Limbic System-Associated Membrane
; TITLE OF INVENTION: Protein and DNA
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,657D
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 317743-102
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 60:
```


Align seg 1/1 to: AAG73854 from: 1 to: 552

1 GCCTTTGTAGTGGACAAAACACCAAAATGGCTACCTCCAAATCCCAGG 50
|||||
179 GlyPheValValAspLysAsnThrLysMetAlaTyrLeuGlnIleProG1 195

51 CATTGCTAAGGTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAAA 100
|||||
195 yileAlaLysValGlyThrTrpLysTyrSerLeuGlnAlaSerSerGlnT 212

101 CCTTGACCTGACTGCTACGCTCCGCTCGCTCAATGCTACCTGCCTCCA 150
|||||
212 hrLeuThrLeuThrValThrSerArgAlaSerAsnAlaThrLeuProPro 228

151 ATTACAGTGACTTCCAAACGAAACAGGACACACAGCAAA 189
|||||
229 IleThrValThrSerLysThrAsnLysAspThrSerLys 241

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAG75614

seq_documentation_block:

ID AAG75614 standard; Protein: 869 AA.

XX AAG75614;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:6378.

XX Human: colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.

XX Homo sapiens.

XX WO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

XX 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI: 2001-235357/24.

XX N-PSDB; AAH35019.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -

XX Claim 11; Page 7851-7854; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.

XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were

CC missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 869 AA;

XX alignment_scores:

XX Quality: 315.00 Length: 63
XX Ratio: 5.000 Gaps: 0
XX Percent Similarity: 100.000 Percent Identity: 100.000

XX alignment_block:

XX US:09-049-696-11 x AAG75614 ..

XX Align seg 1/1 to: AAG75614 from: 1 to: 869

1 GGCCTTTGTAGTGGACAAAACACCAAAATGGCTACCTCCAAATCCCAGG 50
|||||
496 GlyPheValValAspLysAsnThrLysMetAlaTyrLeuGlnIleProG1 512

51 CATTGCTAAGGTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAAA 100
|||||
512 yileAlaLysValGlyThrTrpLysTyrSerLeuGlnAlaSerSerGlnT 529

101 CCTTGACCTGACTGCTACGCTCCGCTCGCTCAATGCTACCTGCCTCCA 150
|||||
529 hrLeuThrLeuThrValThrSerArgAlaSerAsnAlaThrLeuProPro 545

151 ATTACAGTGACTTCCAAACGAAACAGGACACACAGCAAA 189
|||||
546 IleThrValThrSerLysThrAsnLysAspThrSerLys 558

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAB74824

seq_documentation_block:

ID AAB74824 standard; Protein: 914 AA.

XX AAB74824;

XX 13-JUN-2001 (first entry)

XX Human ICACC-1 protein sequence.

XX ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW inflammatory bowel disease; autoimmune disease.

XX Hcmq sapiens.

XX WO9944620-A1.

XX 10-SEP-1999.

XX 03-MAR-1999; 99WO-US04703.

XX 03-MAR-1998; 98US-0076815.

XX (MAGA-) MAGAININ PHARM INC.

XX Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;

XX Nicolaides NC, Zhou Y, Dong Q;

XX WPI: 1999-550979/46;

XX N-PSDB; AAF81927.

XX New nucleic acid encoding calcium activated chloride channel, used to
PT identify, e.g. specific modulators for treating atopic allergy.-

XX Claim 11; Fig 4B; 75pp; English.

XX The present sequence represents the human interleukin 9 (IL-9) induced

CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
 CC have anti-allergic, anti-asthmatic, anti-inflammatory and
 CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
 CC used to alleviate asthma (or more generally atopic allergy), while those
 CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
 CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
 CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
 CC (increased levels) or IBD (reduced levels), also for monitoring
 CC treatment of these conditions. The ICACC proteins can be used:
 CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
 CC reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
 CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
 CC identify modulators and binding partners. ICACC polynucleotides can be
 CC used to generate transgenic animals or recombinant cells, used to screen
 CC for antagonists, also as a source of therapeutic antisense agents or
 CC diagnostic probes (for quantifying mRNA expression, e.g. for
 CC identification of modulators).

XX Sequence 914 AA;

alignment_scores:
 Quality: 315.00 Length: 63
 Ratio: 5.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-11 x AAB74824 ..

Align seg 1/1 to: AAB74824 from: 1 to: 914

1 GGCTTTGTAGTGGACAAAAACACCAAAATGGCGCTACCTCCAAATCCAGG 50
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 541 GlyPheValValAspLysAsnThrLysMetAlaTyrLeuGlnIleProG1 557
 51 CATTGCTAAGGTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAAA 100
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 557 ylleAlaLysValGlyThrTrpLysTyrSerLeuGlnAlaSerSerGlnT 574
 101 CCTTGACCTGACTGTCAGTCCCGTGGCGTCCCAATGCTACCCCTGCCTCCA 150
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 574 hrLeuThrLeuThrValThrSerArgAlaSerAsnAlaThrLeuProPro 590

151 ATTACAGTGACTTCCAAACGACACAGGACACGACAAA 189
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 591 IleThrValThrSerLysThrAsnLysAspThrSerLys 603

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA24514

seq_documentation_block:

ID AAM24514 standard; Protein; 914 AA.

XX AC AAM24514;

XX DT 12-OCT-2001 (first entry)

XX DE C902P predicted amino acid sequence.

XX KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
 KW immunogenic; gene therapy; vaccine; colonic cancer.

XX OS Homo sapiens.

XX PN WO200149716-A2.

XX PD 12-JUL-2001.

XX PF 29-DEC-2000; 2000WO-US35596.

XX PR 30-DEC-1999; 99US-0476296.

PR 10-JAN-2000; 2000US-0480321.

PR 15-FEB-2000; 2000US-0504629.

PR 06-MAR-2000; 2000US-0519444.

PR 19-MAY-2000; 2000US-0575251.
 PR 29-JUN-2000; 2000US-0609448.
 PR 28-AUG-2000; 2000US-0649811.
 XX
 PA (CORI-) CORIXA CORP.

XX Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
 PI King GE, Wang T, Jiang Y;
 XX WPI; 2001-441847/47.

XX Colon tumor associated proteins and nucleic acids useful for the
 PT prevention, diagnosis and treatment of colonic cancer -

PS Claim 2; Page 440-443; 472pp; English.

XX The present invention describes colon tumour associated proteins (I) and
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumour associated protein (TCAP)
 CC expression, such as colonic cancer. For example, (I) and (II) may be
 CC used to treat disorders associated with decreased expression by
 CC rectifying mutations or deletions in a patient's genome that affect the
 CC activity of TCAPs by expressing inactive proteins or to supplement the
 CC patients own production of them. Additionally, (II) may be used to
 CC produce the TCAP proteins, by inserting the nucleic acids into a host
 CC cell culturing the cell to express the protein. (II) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and
 CC quantitate the presence of similar nucleic acids in samples, and
 CC therefore which patients may be in need of restorative therapy. (I) may
 CC also be used as antigens in the production of antibodies against TCAPs
 CC and in assays to identify modulators of TCAP expression and activity.
 CC Anti-(I) antibodies and antagonists may also be used to down regulate
 CC TCAP expression and activity. The anti-(I) antibodies may also be used
 CC as diagnostic agents for detecting the presence of TCAPs in samples
 CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512
 CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
 CC given in the exemplification of the present invention.

XX Sequence 914 AA;

alignment_scores:
 Quality: 315.00 Length: 63
 Ratio: 5.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-11 x AAM24514 ..

Align seg 1/1 to: AAM24514 from: 1 to: 914

1 GGCTTTGTAGTGGACAAAAACACCAAAATGGCGCTACCTCCAAATCCAGG 50
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 541 GlyPheValValAspLysAsnThrLysMetAlaTyrLeuGlnIleProG1 557
 51 CATTGCTAAGGTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAAA 100
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 557 ylleAlaLysValGlyThrTrpLysTyrSerLeuGlnAlaSerSerGlnT 574
 101 CCTTGACCTGACTGTCAGTCCCGTGGCGTCCCAATGCTACCCCTGCCTCCA 150
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 574 hrLeuThrLeuThrValThrSerArgAlaSerAsnAlaThrLeuProPro 590
 151 ATTACAGTGACTTCCAAACGACACAGGACACGACAAA 189
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 591 IleThrValThrSerLysThrAsnLysAspThrSerLys 603

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB73716

seq_documentation_block:

```

ID XX AAB73716 standard; Protein; 914 AA.
AC
XX AAB73716;
XX
XX 11-SEP-2001 (first entry)
DT
XX Human CLC1 protein, SEQ ID NO:2.
DE
XX
XX Human CLC1; goblet cell; mouse Gob-5 orthologue; drug screening;
KW expression inhibition; antisense therapy; gene therapy;
KW chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic.
XX
XX Homo sapiens.
XX
XX KEY Location/Qualifiers
FH Misc-difference 152
FT /note= "Encoded by AGG in AAH46124"
XX
XX WO200138530-A1.
XX 31-MAY-2001.
XX
XX 22-NOV-2000; 2000WO-JP08232.
XX
XX 24-NOV-1999; 99JP-0333479.
XX 27-APR-2000; 2000JP-0127589.
XX
XX (TAKE ) TAKEDA CHEM IND LTD.
XX
XX Nakanishi A, Morita S;
XX WPI; 2001-355935/37.
XX DR N-PSDB; AAH46102, AAH46124.
XX
XX New antisense nucleotide, useful for treatment and prevention of
XX bronchial asthma and chronic obstructive pulmonary disease
XX
XX Claim 2; Page 76-80; 104pp; Japanese.
XX
XX The invention relates to an antisense nucleotide targetted to the mouse
XX Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
XX the CLC1 gene (coding sequence shown in AAH46102). The invention also
XX relates to an antibody specific for the Gob-5 protein, medical and
XX diagnostic compositions containing the antisense nucleotide or the
XX antibody, and methods and kits for screening for compounds which inhibit
XX the protein. Gob-5 and CLC1 are proteins expressed by goblet cells.
XX The antisense oligonucleotides and antibody are therefore useful for the
XX treatment and prevention of bronchial asthma and chronic obstructive
XX pulmonary disease. The present sequence represents human CLC1 protein.
XX
XX Sequence 914 AA;
SQ

alignment_scores:
Quality: 315.00 Length: 63
Ratio: 5.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-11 x AAB73716 ..
Align seg 1/1 to: AAB73716 from: 1 to: 914

1 GGCTTTGAGTGACAAAACACCAAAATGGCTACCTCCAATCCAGG 50
|||||
541 GlyPheValValAspLysAsnThrLysMetAlaTyrLeuGlnIleProG1 557
|||||
51 CATTGCTAGGTGGCAGCTTGGAAATACAGTCTGCAAGCAAGCTCACAA 100
|||||
557 YileAlaLysValGlyThrIrrpLysTyrSerLeuGlnAlaSerSerGlnT 574
|||||
101 CCTTGACCTGACTGTGACGTCGCCGTCGCTCAATGCTACCTGCCTCCA 150
|||||

seq name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAB74733
documentation_block:
ID AAB74733 standard; Protein; 914 AA.
AC AAB74733;
XX
XX 12-JUN-2001 (first entry)
DT
XX Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.
DE
XX
XX Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; neurotropic; anticonvulsant; vaccine;
KW antialzheimer; antiparkinsonian; antimicrobial; vulnerary; gene therapy;
KW immune disorder; hyperproliferative disorder; cardiovascular disease;
KW cancer; angiogenic disorder; neurological disorder; infectious disease;
KW wound healing; regeneration; chemotaxis; chromosome 1.
XX
XX Homo sapiens.
XX
XX WO200112775-A2.
XX 22-FEB-2001.
XX
XX 16-AUG-2000; 2000WO-US22325.
XX
XX 17-AUG-1999; 99US-0149182.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ni J, Florence KA, Fiscella M, Wei P, Baker KP;
XX Birse CB, Young PE, Komatsoulis GA, Moore PA, Soppet DR;
XX WPI; 2001-147550/15.
XX N-PSDB; AAF81787.
XX
XX Nucleic acids encoding 25 human secreted polypeptides, useful for
XX preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX disease and diabetic retinopathy
XX
XX Claim 11; Page 459-460; 485pp; English.
XX
XX AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733
XX to AAB74772. Human secreted proteins can have activities based on the
XX tissues and cells they are expressed in. Example of activities include:
XX immunomodulatory; antisclerotic; dermatological; immunosuppressive;
XX antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;
XX vascular; anti-angiogenic; ophthalmological; neuroprotectant; neurotropic;
XX anticonvulsant; antialzheimers; antiparkinsonian; antimicrobial; and
XX vulnerary. Human secreted proteins can be used in gene therapy and
XX vaccine. Human secreted protein nucleotide sequences (NAMI) and proteins
XX (tEPI) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate polypeptide expression. For example, NAMI
XX and tEPI may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patients genome
XX that affect the activity of proteins by expressing inactive proteins or
XX to supplement the patients own production of polypeptides. Disorders that
XX may be prevented, diagnosed and/or treated include immune disorders,
XX hyperproliferative disorders (e.g. cancers), cardiovascular diseases,
XX angiogenic disorders, neurological disorders, infectious diseases and/or
XX for promoting wound healing, regeneration and /or chemotaxis. AAF81778 to
XX AAF81786 and AAB74732 represent sequences used in the exemplification of
XX the present invention.
XX
XX Sequence 914 AA;
SQ
```

CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX
SQ Sequence 925 AA;

alignment_scores:
Quality: 315.00 Length: 63
Ratio: 5.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-11 x AAG75474 ..

Align seg 1/1 to: AAG75474 from: 1 to: 925

1 GGCTTTGCTAGTGACAAAACACCAAAATGGCTACTCTCCAAATCCGAGG 50
|||||
552 GlyPheValValAspLysAsnThrLysMetAlaTyrLeuGlnIleProG1 568

51 CATTGCTAAGGTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAAA 100
|||||
568 YileAlaLysValGlyThrTrpLysTyrSerLeuGlnAlaSerSergInt 585

101 CCTTGACCCCTGACTGTACGTCGCCGTCGCTCAATGCTACCCCTCCCA 150
|||||
585 hrLeuThrLeuThrValThrSerArgAlaSerAsnAlaThrLeuProPro 601

151 ATTACAGTGACTTCCAAAGCAACAGGACACGACGACAAA 189
|||||
602 IleThrValThrSerLysThrAsnLysAspThrSerLys 614

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAH74822

seq_documentation_block:
ID AAB74822 standard; Protein: 913 AA.
XX
AC AAB74822;
XX
DT 13-JUN-2001 (first entry)
XX
DE Murine ICACC-1 protein sequence.
XX
KW ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW inflammatory bowel disease; autoimmune disease.
OS Mus sp.
XX
PN WO9944620-A1.
XX
PD 10-SEP-1999.
XX
PF 03-MAR-1999; 99WO-US04703.
XX
PR 03-MAR-1998; 98US-0076815.
XX
PA (MAGA-) MAGATIN PHARM INC.
XX
PI Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;
PI Nicolaides NC, Zhou Y, Dong Q;
XX

CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX
SQ Sequence 925 AA;

alignment_scores:
Quality: 315.00 Length: 63
Ratio: 5.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-11 x AAB74733 ..

Align seg 1/1 to: AAB74733 from: 1 to: 914

1 GGCTTTGCTAGTGACAAAACACCAAAATGGCTACTCTCCAAATCCGAGG 50
|||||
541 GlyPheValValAspLysAsnThrLysMetAlaTyrLeuGlnIleProG1 557

51 CATTGCTAAGGTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAAA 100
|||||
557 YileAlaLysValGlyThrTrpLysTyrSerLeuGlnAlaSerSergInt 574

101 CCTTGACCCCTGACTGTACGTCGCCGTCGCTCAATGCTACCCCTCCCA 150
|||||
574 hrLeuThrLeuThrValThrSerArgAlaSerAsnAlaThrLeuProPro 590

151 ATTACAGTGACTTCCAAAGCAACAGGACACGACGACAAA 189
|||||
591 IleThrValThrSerLysThrAsnLysAspThrSerLys 603

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAH75474

seq_documentation_block:
ID AAG75474 standard; Protein: 925 AA.
XX
AC AAG75474;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6238.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 1.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR N-PSDB; AAH34879.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 7686-7690; 9803pp; English.

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions

pharmaceutical; receptor immunoadhesin; gene mapping.

KW
XX
OS
XX
PN
XX
XX
PD
XX
XX
PF
XXHomo sapiens.
WO9903088-A2.
09-DEC-1999.

02-JUN-1999; 99WO-US12252.

02-JUN-1998; 98US-0087607.

02-JUN-1998; 98US-0087609.

02-JUN-1998; 98US-0087759.

03-JUN-1998; 98US-0087827.

04-JUN-1998; 98US-0088021.

04-JUN-1998; 98US-0088025.

04-JUN-1998; 98US-0088028.

04-JUN-1998; 98US-0088029.

04-JUN-1998; 98US-0088030.

04-JUN-1998; 98US-0088033.

05-JUN-1998; 98US-0088326.

05-JUN-1998; 98US-0088167.

05-JUN-1998; 98US-0088202.

05-JUN-1998; 98US-0088212.

05-JUN-1998; 98US-0088217.

09-JUN-1998; 98US-0088655.

10-JUN-1998; 98US-0088722.

10-JUN-1998; 98US-0088730.

10-JUN-1998; 98US-0088734.

10-JUN-1998; 98US-0088740.

10-JUN-1998; 98US-0088741.

10-JUN-1998; 98US-0088742.

10-JUN-1998; 98US-0088810.

10-JUN-1998; 98US-0088811.

10-JUN-1998; 98US-0088824.

10-JUN-1998; 98US-0088825.

10-JUN-1998; 98US-0088826.

11-JUN-1998; 98US-0088858.

11-JUN-1998; 98US-0088861.

11-JUN-1998; 98US-0088863.

11-JUN-1998; 98US-0088876.

12-JUN-1998; 98US-0089090.

12-JUN-1998; 98US-0089105.

16-JUN-1998; 98US-0089440.

16-JUN-1998; 98US-0089512.

16-JUN-1998; 98US-0089514.

17-JUN-1998; 98US-0089532.

17-JUN-1998; 98US-0089538.

17-JUN-1998; 98US-0089598.

17-JUN-1998; 98US-0089599.

17-JUN-1998; 98US-0089600.

17-JUN-1998; 98US-0089603.

17-JUN-1998; 98US-0089653.

18-JUN-1998; 98US-0089801.

18-JUN-1998; 98US-0089907.

18-JUN-1998; 98US-0089908.

19-JUN-1998; 98US-0089947.

19-JUN-1998; 98US-0089948.

19-JUN-1998; 98US-0089952.

22-JUN-1998; 98US-0090246.

22-JUN-1998; 98US-0090252.

22-JUN-1998; 98US-0090254.

23-JUN-1998; 98US-0090349.

23-JUN-1998; 98US-0090355.

24-JUN-1998; 98US-0090429.

24-JUN-1998; 98US-0090431.

24-JUN-1998; 98US-0090435.

24-JUN-1998; 98US-0090444.

24-JUN-1998; 98US-0090445.

24-JUN-1998; 98US-0090461.

24-JUN-1998; 98US-0090472.

24-JUN-1998; 98US-0090535.

24-JUN-1998; 98US-0090538.

PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 02-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 24-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097218.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 31-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 98US-0115565.

```
XX (GETH ) GENENTECH INC.
XX Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
XX Wood WI, Yuan J;
XX WPI; 2000-072883/06.
XX N-PSDB; AAZ65095.
XX Membrane-bound proteins and related nucleotide sequences
XX claim 12; Fig 274; 822pp; English.
XX The invention provides membrane-bound PRO polypeptides and
XX polynucleotides encoding them. The PRO sequences of the invention were
XX identified based on extracellular domain homology screening. The PRO
XX sequences have homology with proteins including LDL receptors, TIE
XX ligands and various enzymes. The membrane-bound proteins and receptor
XX molecules are useful as pharmaceutical and diagnostic agents. Receptor
XX immunoadhesins, for instance, can be used as therapeutic agents to block
XX receptor-ligand interactions. The membrane-bound proteins can also be
XX employed for screening of potential peptide or small molecule inhibitors
XX of the relevant receptor/ligand interaction. The PRO encoding sequences
XX are useful as hybridization probes, in chromosome and gene mapping and in
XX the generation of antisense RNA and DNA. PRO nucleic acid sequences
XX will also be useful for the preparation of PRO polypeptides, especially
XX by recombinant techniques.
XX Sequence 919 AA;

alignment_scores:
  Quality: 207.00      Length: 63
  Ratio: 3.632        Gaps: 1
  Percent Similarity: 90.476  Percent Identity: 63.492

alignment_block:
US-09-049-696-11 x AAY66749 ..
  Align seg 1/1 to: AAY66749 from: 1 to: 919
    4 TTTGTAGTGGACAAACACCAAAATGGCTTACCTCCAAATCCAGGCAT 53
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    543 PheThrValAspAlaThrSerLysMetAlaTyrLeuSerIleProGlyTh 559
    54 TGCTAAGTTGGCAGTCTGGAATACAGTCTGCAAGCAAGCTCA.....C 97
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    559 rAlaLysValGlyThrTrpAlaTyrAsnLeuGlnAlaLysAlaAsnProG 576
    98 AAACCTTGACCTGACTGTCCAGTCCCGTGGCTGCTCAATGCTACCTGCCT 147
      :|||: ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
    576 luThrLeuThrIleThrValThrSerArgAlaAlaAsnSerSerValPro 592
    148 CCAATTACAGTGCAGTCTCCAAACGACCAAGGACACACCAGC 186
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    593 ProIleThrValAsnAlaLysMetAsnLysAspValAsn 605

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB87560

seq_documentation_block:
ID AAB87560 standard; Protein; 919 AA.
XX
XX AAB87560;
AC
XX
XX 15-MAY-2001 (first entry)
XX
XX
XX Human PRO1124.
XX
XX Human; PRO protein; mapping.
XX
XX Homo sapiens.
XX
XX WO200116318-A2.
PN
```

```
XX 08-MAR-2001.
XX
XX 24-AUG-2000; 2000WO-US23328.
XX
XX 01-SEP-1999; 99WO-US20111.
XX 15-SEP-1999; 99WO-US21090.
XX 07-DEC-1999; 99US-0169495.
XX 09-DEC-1999; 99US-0170262.
XX 11-JAN-2000; 2000US-0175481.
XX 18-FEB-2000; 2000WO-US04341.
XX 18-FEB-2000; 2000WO-US04342.
XX 22-FEB-2000; 2000WO-US04414.
XX 01-MAR-2000; 2000WO-US05601.
XX 03-MAR-2000; 2000US-0187202.
XX 25-APR-2000; 2000US-0199397.
XX 22-MAY-2000; 2000WO-US14042.
XX 05-JUN-2000; 2000US-0209832.
XX (GETH ) GENENTECH INC.
XX Eåton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
XX Grimadi CJ, Gurney AL, Watanabe CK, Wood WI;
XX WPI; 2001-183260/18.
XX N-PSDB; AAF92092.
XX Eighty four nucleic acids encoding PRO polypeptides, useful in
XX molecular biology, including use as hybridization probes, and in
XX chromosome and gene mapping.
XX Claim 12; Fig. 70; 278pp; English.
XX The present sequence is a human PRO polypeptide (secreted and
XX transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
XX anti-PRO antibodies are useful for preparation of a medicament useful in
XX the treatment of a condition which is responsive to the PRO protein,
XX agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
XX employed as molecular weight markers for protein electrophoresis. The PRO
XX coding sequence has applications in molecular biology, including use as
XX hybridisation probes, and in chromosome and gene mapping.
XX Sequence 919 AA;

alignment_scores:
  Quality: 207.00      Length: 63
  Ratio: 3.632        Gaps: 1
  Percent Similarity: 90.476  Percent Identity: 63.492

alignment_block:
US-09-049-696-11 x AAB87560 ..
  Align seg 1/1 to: AAB87560 from: 1 to: 919
    4 TTTGTAGTGGACAAACACCAAAATGGCTTACCTCCAAATCCAGGCAT 53
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    543 PheThrValAspAlaThrSerLysMetAlaTyrLeuSerIleProGlyTh 559
    54 TGCTAAGTTGGCAGTCTGGAATACAGTCTGCAAGCAAGCTCA.....C 97
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    559 rAlaLysValGlyThrTrpAlaTyrAsnLeuGlnAlaLysAlaAsnProG 576
    98 AAACCTTGACCTGACTGTCCAGTCCCGTGGCTGCTCAATGCTACCTGCCT 147
      :|||: ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||:
    576 luThrLeuThrIleThrValThrSerArgAlaAlaAsnSerSerValPro 592
    148 CCAATTACAGTGCAGTCTCCAAACGACCAAGGACACACCAGC 186
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    593 ProIleThrValAsnAlaLysMetAsnLysAspValAsn 605

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB65272
```

```
seq_documentation_block:
ID  AAB65272 standard; Protein; 919 AA.
AC  AAB65272;
XX
XX
XX  02-APR-2001 (first entry)
XX
XX  Human PRO1124 (UNQ562) protein sequence SEQ ID NO:379.
XX
XX  Human; secreted and transmembrane protein; PRO; cytosstatic;
KW  cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW  diagnostic assay.
XX
XX  Homo sapiens.
OS
XX  WO2000073454-A1.
XX
XX  07-DEC-2000.
XX
XX  30-MAR-2000; 2000WO-US08439.
XX
XX  02-JUN-1999; 99WO-US12252.
XX  23-JUN-1999; 99US-0141037.
XX  07-JUL-1999; 99US-0143048.
XX  20-JUL-1999; 99US-0144758.
XX  26-JUL-1999; 99US-0145698.
XX  28-JUL-1999; 99US-0146222.
XX  17-AUG-1999; 99US-0149396.
XX  15-SEP-1999; 99WO-US21090.
XX  15-SEP-1999; 99WO-US21547.
XX  08-OCT-1999; 99US-0158663.
XX  30-NOV-1999; 99WO-US28313.
XX  01-DEC-1999; 99WO-US28301.
XX  16-DEC-1999; 99WO-US30095.
XX  20-DEC-1999; 99WO-US30911.
XX  05-JAN-2000; 2000WO-US00219.
XX  06-JAN-2000; 2000WO-US00376.
XX  11-FEB-2000; 2000WO-US03565.
XX  18-FEB-2000; 2000WO-US04341.
XX  22-FEB-2000; 2000WO-US04414.
XX  24-FEB-2000; 2000WO-US04914.
XX  02-MAR-2000; 2000WO-US05004.
XX  02-MAR-2000; 2000WO-US05841.
XX  15-MAR-2000; 2000WO-US06884.
XX  20-MAR-2000; 2000WO-US07377.
XX
XX  (GETH ) GENENTECH INC.
XX
XX  Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI  Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
PI  Grimaldi CJ, Gurney AL, Kljavin IJ, Napier MA, Pan J, Paoni NF;
PI  Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood W;
PI  Zhang Z;
XX
XX  WPI; 2001-032160/04.
XX  N-PSDB; AAF44241.
XX
XX  PRO polynucleotides used to produce polypeptides used to target
XX  bioactive molecules such as toxins, radiolabels or antibodies, to
XX  specific cells, to cause targeted cell death -
XX
XX  Claim 12; Fig 274; 935pp; English.
XX
XX  The present invention describes human secreted and transmembrane PRO
XX  proteins. The PRO proteins have cytostatic activity. The PRO proteins
XX  can be used for targeted delivery of bioactive molecules, such as
XX  toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
XX  sequences, and their fragments, can be used as hybridisation probes, in
XX  chromosomal and gene mapping, and in the generation of anti-sense RNA
XX  and DNA. They may also be used to produce transgenic animals which are
XX  used to develop and screen therapeutically useful reagents. The PRO
XX  nucleotide and protein sequence can be used for tissue typing and in
XX  treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
XX
CC  AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
CC  in the isolation of human PRO sequences. AAF44087 to AAF44269 and
CC  AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC  sequences given in the exemplification of the present invention.
XX
XX  Sequence 919 AA;
XX
XX  alignment_scores:
XX      Quality: 207.00      Length: 63
XX      Ratio: 3.632      Gaps: 1
XX      Percent Similarity: 90.476      Percent Identity: 63.492
XX
XX  alignment_block:
XX  US-09-049-696-11 x AAB65272
XX
XX  Align seg 1/1 to: AAB65272 from: 1 to: 919
XX
XX  4 TTTGTAGTGGACAAACACCAAAATGGCTACTCTCAATCCAGGCAT 53
XX  |||:||||| :||:|||||:|||||:|||||:|||||:|||||
XX  543 PheThrValAspAlaThrSerLysMetAlaTyrLeuSerIleProGlyTh 559
XX  54 TGCTAAGTTGGCAGCTTGGAAATACAGTCTGCAAGCAAGCTCA.....C 97
XX  |||:|||||:|||||:|||||:|||||:|||||:|||||
XX  559 rAlaLysValGlyThrTrpAlaTyrAsnLeuGlnAlaLysAlaAsnProG 576
XX  98 AACCTTGACCTGACTGTACGTCCGTCGCTGCTCAATGCTACCGCTGCCT 147
XX  :|||:|||||:|||||:|||||:|||||:|||||:|||||
XX  576 luThrLeuThrIleThrValThrSerArgAlaAlaAsnSerSerValPro 592
XX  148 CCAATTACAGTGACTTCCAAAACGACACAGGACACCAGC 186
XX  |||:|||||:|||||:|||||:|||||:|||||:|||||
XX  593 ProIleThrValAsnAlaLysMetAsnLysAspValAsn 605
XX
XX  seq_name: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT: AAB53310
XX
XX  seq_documentation_block:
XX  ID  AAB53310 standard; Protein; 218 AA.
XX
XX  AC  AAB53310;
XX
XX  DT  09-MAR-2001 (first entry)
XX
XX  DE  Human colon cancer antigen protein sequence SEQ ID NO:850.
XX
XX  KW  Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX  identification; cytostatic; cardiocactive; neuroprotective; vulnerary;
XX  immunomodulatory; muscular; gynaecological; gastrointestinal;
XX  nephrotropic; antiinfective; antibacterial; gene therapy; wound;
XX  neural disorder; immune system disorder; muscular disorder;
XX  reproductive disorder; gastrointestinal disorder; renal disorder;
XX  infectious disease; cardiovascular disorder.
XX
XX  OS  Homo sapiens.
XX
XX  PN  WO2000055351-A1.
XX
XX  PD  21-SEP-2000.
XX
XX  PF  08-MAR-2000; 2000WO-US05883.
XX
XX  PR  12-MAR-1999; 99US-0124270.
XX
XX  PA  (HUMA-) HUMAN GENOME SCI INC.
XX
XX  PI  Rosen CA, Ruben SM;
XX  i
XX  DR  WPI; 2000-587534/55.
XX  DR  N-PSDB; AAC98067.
XX
XX  PT  Colon cancer associated gene sequences, referred to as colon cancer
XX  antigens, useful for the treatment, prevention, and diagnosis of colon
XX  disorders such as colon cancer -
```


This Page Blank (uspto)

OM of: US-09-049-696-10 to: SPTRMBL_17:* out_format : pfs
Date: Mar 30, 2002 2:46 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+np2.model -DEV=slp
-Q/cgn2_1/USPTO_spool/US09049696/runat_28032002_145238_2053/app_query.fasta_1.12579
-DB=SPTRMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-FGAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsub2
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1_705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPPY
-WAIT -THREADS=1

Search information block:
Query: US-09-049-696-10
Query length: 229
Database: SPTRMBL_17:*
Database sequences: 473505
Database length: 146272329
Search time (sec): 805.760000

score_list:	Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
	sp_human:Q05151	+	396.00	863.90	8.3e-41	914	! 095151 homo sapiens (human)..ca
	sp_human:Q9UPC6	+	396.00	863.90	8.3e-41	914	! 09upc6 homo sapiens (human)..ca
	sp_human:Q9UNF6	+	396.00	863.90	8.3e-41	914	! 09unf6 homo sapiens (human)..ca
	sp_rdent:Q88826	+	319.00	691.05	3.5e-31	913	! 088826 mus musculus (mouse)..gc
	sp_rdent:Q9D776	+	319.00	691.05	3.5e-31	913	! 09d776 mus musculus (mouse)..ch
	sp_human:Q9NXP1	+	263.00	570.75	3.5e-24	469	! 09npx1 homo sapiens (human)..cd
	sp_human:Q9UNF7	+	263.00	569.30	3.6e-24	917	! 09unf7 homo sapiens (human)..ca
	sp_mammal:Q9TUB5	+	244.00	522.65	8.4e-22	917	! 09tub5 sus scrofa (pig).epithe
	sp_mammal:Q18742	+	219.00	467.70	1.1e-18	794	! 018742 bos taurus (bovine)..lu
	sp_mammal:Q18743	+	219.00	467.43	1.1e-18	820	! 018743 bos taurus (bovine)..lu
	sp_mammal:Q18741	+	219.00	466.63	1.1e-18	905	! 018741 bos taurus (bovine)..lu
	sp_rdent:Q9R070	+	201.00	426.25	2.0e-16	902	! 09r070 mus musculus (mouse)..ca
	sp_rdent:Q9EQR4	+	196.00	415.03	8.5e-16	902	! 09edr4 mus musculus (mouse)..en
	sp_rdent:Q88860	+	193.00	408.30	2.0e-15	901	! 088860 mus musculus (mouse)..ch
	sp_rdent:Q9QX15	+	193.00	408.29	2.0e-15	902	! 09qx15 mus musculus (mouse)..ca
	sp_human:Q9YQC9	+	171.00	358.54	1.1e-12	943	! 09uqc9 homo sapiens (human)..ch
	sp_human:Q9Y6N2	+	171.00	358.54	1.1e-12	943	! 09y6n2 homo sapiens (human)..ca
	sp_human:Q9P2W6	+	74.50	157.89	1.22	132	! 09p2w6 homo sapiens (human)..ch
	sp_plant:Q9W332	+	73.00	143.00	1.99	545	! 09m332 arabidopsis thaliana (mc
	sp_vertbrate:Q9W7R4	-	71.00	125.84	3.78	2590	! 09w7r4 brachydanio rerio (zebr
	sp_vertbrate:Q9IBG9	-	70.50	128.05	4.29	1718	! 09ibg9 oryzias latipes (medaka
	sp_rdent:Q9EPW5	-	68.00	128.71	8.53	795	! 09epw5 mus musculus (mouse)..tc
	sp_rdent:Q9EP01	-	68.00	128.71	8.53	795	! 09epq1 mus musculus (mouse)..tc
	sp_rdent:Q9WQ04	-	68.00	128.59	8.54	806	! 09wq04 mus musculus (mouse)..tl
	sp_rdent:Q9EPW9	-	68.00	128.59	8.54	806	! 09epw9 mus musculus (mouse)..tc
	sp_human:Q9N9V7	-	68.00	127.14	8.60	964	! 09nv77 homo sapiens (human)..cd
	sp_human:Q9N9V1	-	68.00	126.48	8.63	1045	! 09nvw1 homo sapiens (human)..cd
	sp_rdent:Q9JLC1	-	68.00	119.91	8.93	2346	! 09jlc1 mus musculus (mouse)..c
	sp_rdent:Q9W5S6	-	68.00	118.72	8.99	2715	! 09wts6 mus musculus (mouse)..t
	sp_rdent:Q9CTQ9	+	67.50	133.46	9.56	386	! 09cyq9 mus musculus (mouse)..28
	sp_invertebrate:Q9V123	+	67.50	123.71	10.06	1280	! 09v123 drosophila melanogast
	sp_human:Q9UPL3	-	67.00	117.89	11.90	2281	! 09upl3 homo sapiens (human)..r
	sp_human:Q9Y6L8	-	67.00	117.75	11.91	2321	! 09y6l8 homo sapiens (human)..r
	sp_human:Q9UM47	-	67.00	117.75	11.91	2321	! 09um47 homo sapiens (human)..r
	sp_mammal:Q77780	-	66.00	124.74	15.14	745	! 077780 bos taurus (bovine)..fed
	sp_bacteria:Q9L950	+	66.00	116.14	15.83	2147	! 09l950 pseudomonas putida. out
	sp_mammal:Q28733	+	65.50	105.56	19.21	6875	! 028733 oryctolagus cuniculus (
	sp_plant:Q9WAG9	-	65.00	124.55	19.97	579	! 09wag9 arabidopsis thaliana (mc
	sp_vertbrate:Q73634	+	65.00	122.72	20.16	725	! 073634 xenopus laevis (afrikan
	sp_fungi:Q9UVG6	+	65.00	117.73	20.70	1340	! 09uvg6 pichia pastoris (yeast)

sp_human:Q10466 + 65.00 93.34 23.50 26926 ! Q10466 homo sapiens (human
sp_plant:Q38887 - 64.50 121.68 23.28 718 ! Q38887 arabidopsis thaliana
sp_plant:Q9SVS2 - 64.50 121.68 23.28 718 ! Q9SVS2 arabidopsis thaliana
sp_invertebrate:Q9U8J9 + 64.50 121.16 23.34 765 ! Q9u8j9 neospora caninum. t
sp_bacteria:Q51133 + 64.00 120.85 26.84 692 ! Q51133 neisseria meningitidi
seq_name: sp_human:Q95151
seq_documentation_block:
ID Q95151 PRELIMINARY; PRT; 914 AA.
AC Q95151;
DC 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DD DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
OS HCLCAL.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Elble R.C., Ji H.L., Schreur K.D., Fuller C.M.,
PA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
of human CLCAL, the first human member of the family of Ca2+-activated
CL-channel proteins.";
RL Genomics 54:200-214(1998).
DR EMBL: AF039400; AAC95428.1;
DR InterPro: IPR000131; ATPase_gamma.
DR InterPro: IPR002035; VWFA.
DR PROSITE: PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE: PS0234; VWFA; 1.
DR SMART: SM00327; VWA; 1.
SQ SEQUENCE 914 AA; 100174 MW; 7066440C46557FA3 CRC64;

alignment_scores:
Quality: 396.00 Length: 76
Ratio: 5.211 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-10 x Q95151
Align seg 1/1 to: Q95151 from: 1 to: 914

1 GCACAGTATCGTGGACAGCCGTCGGAAGGACACTTGTCTTAT 50
|||||
504 GlyThrValIleValAspSerThrValGlyLysAspThrLeuPheLeu1 520
51 CACCTGGACAACGACGCTCCCAATCTCTCTGGATCCAGTGGAC 100
|||||
520 eThrTrpThrGlnProGlnIleLeuLeuTrpAspProSerGlyG 537
101 AGAACCAAGTGGCTTTGTAGTGGACAAACACCAAAATGGCTACCTC 150
|||||
537 InLysInGlyGlyPheValValAspLysAsnThrLysMetAlaTrLeu 553
141 CAATCCAGGATGCTAAGTGGCATTGGAAATACAGTCTGCAAGC 200
|||||
354 GlnIleProGlyIleAlaLysValGlyThrLysTyrSerLeuGlnAl 570
201 AAGCTCACAACTTGACCTGACTGTC 228
|||||
570 aSerSerGlnThrLeuThrLeuThrVal 579
seq_name: sp_human:Q9UPC6
seq_documentation_block:
ID Q9UPC6 PRELIMINARY; PRT; 914 AA.

Q9UPC6;
 AC 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
 GN HCLCAL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99047526; PubMed=9828122;
 RA Gruber A.D., Elble R.C., Ji H.L., Schreier K.D., Fuller C.M.,
 RA Pauli B.U.;
 RT "Genomic cloning, molecular characterization, and functional analysis
 of human CLCAL, the first human member of the family of Ca²⁺-activated
 Cl-channel proteins."
 RL Genomics 54:200-214(1998).
 DR EMBL; AF039401; AAC95429.1; -
 DR InterPro; IPR000131; ATPase_gamma.
 DR InterPro; IPR002035; vWFA.
 DR PROSITE; PS00153; ATPase_GAMMA; UNKNOWN_1.
 DR PROSITE; PS0234; vWFA; 1.
 DR SMART; SM00327; vWA; 1.
 DR SEQUENCE 914 AA; 100160 MW; 7761362B66752AA3 CRC64;
 SQ
 alignment_scores:
 Quality: 396.00 Length: 76
 Ratio: 5.211 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-049-696-10 x Q9UPC6 ..
 Align seg 1/1 to: Q9UPC6 from: 1 to: 914
 1 GGACAGTGTGTCGGACAGCAGCCGTCGGAAAGGACACTTTGTTCTTAT 50
 504 GlyThrValIleValAspSerThrValGlyLysAspThrLeuPheLeuI 520
 51 CACCTGGACACACCGCCCTCCCAATCTCTGGGATCCAGTGGAC 100
 520 eThrTrpThrGlnProGlnIleLeuLeuTrpAspProSerGlyG 537
 101 AGAAGCAAGTGGCTTTGTAGTGGACAAACACCAAAATGGCCTACCTC 150
 537 InLysGlnGlyGlyPheValValAspLysAsnThrLysMetAlaTrLeu 553
 151 CAAATCCAGGCAATGTAGTGGACAAACACCAAAATGGCCTACCTC 200
 554 GlnIleProGlyIleAlaLysValGlyThrTrpLysTyrSerLeuGlnAl 570
 201 AAGCTCACAACTTGACCTGACTGTC 228
 570 aSerSerGlnThrLeuThrLeuThrVal 579
 seq_name: sp_human:Q9UPC6
 seq_documentation_block:
 ID Q9UPC6 PRELIMINARY; PRT; 914 AA.
 AC Q9UPC6;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 1.
 GN CACCL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

SEQUENCE FROM N.A.
 RC TISSUE=SMALL INTESTINE, AND COLON;
 RX MEDLINE=99364503; PubMed=10437792;
 RA Agnel M., Vermet T., Culouscou J.M.;
 RT "Identification of three novel members of the calcium-dependent
 chloride channel (CaCC) family predominantly expressed in the
 digestive tract and trachea."
 RL FEBS Lett. 455:295-301(1999).
 DR EMBL; AF127036; AAD25487.1; -
 DR InterPro; IPR000131; ATPase_gamma.
 DR InterPro; IPR002035; vWFA.
 DR SMART; SM00327; vWA; 1.
 DR PROSITE; PS00153; ATPase_GAMMA; UNKNOWN_1.
 DR PROSITE; PS0234; vWFA; 1.
 DR SEQUENCE 914 AA; 100202 MW; B58B451D97887D6B CRC64;
 SQ
 alignment_scores:
 Quality: 396.00 Length: 76
 Ratio: 5.211 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-049-696-10 x Q9UNF6 ..
 Align seg 1/1 to: Q9UNF6 from: 1 to: 914
 1 GGACAGTGTGTCGGACAGCAGCCGTCGGAAAGGACACTTTGTTCTTAT 50
 504 GlyThrValIleValAspSerThrValGlyLysAspThrLeuPheLeuI 520
 51 CACCTGGACACACCGCCCTCCCAATCTCTGGGATCCAGTGGAC 100
 520 eThrTrpThrGlnProGlnIleLeuLeuTrpAspProSerGlyG 537
 101 AGAAGCAAGTGGCTTTGTAGTGGACAAACACCAAAATGGCCTACCTC 150
 537 InLysGlnGlyGlyPheValValAspLysAsnThrLysMetAlaTrLeu 553
 151 CAAATCCAGGCAATGTAGTGGACAAACACCAAAATGGCCTACCTC 200
 554 GlnIleProGlyIleAlaLysValGlyThrTrpLysTyrSerLeuGlnAl 570
 201 AAGCTCACAACTTGACCTGACTGTC 228
 570 aSerSerGlnThrLeuThrLeuThrVal 579
 seq_name: sp_rodent:O88826
 seq_documentation_block:
 ID O88826 PRELIMINARY; PRT; 913 AA.
 AC O88826;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE GOR-5 PROTEIN.
 GN CLCA3 OR GOR-5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=INTESTINE;
 RA Komiya T., Tanigawa Y., Hirohashi S.;
 RT "Cloning of a gene, gob-5, which is expressed in intestinal goblet
 cells in mice."
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 DE EMBL; AB017156; BAA33743.1; -
 DR MGD; MGI:1346342; Clca3.
 DR InterPro; IPR000131; ATPase_gamma.
 DR InterPro; IPR002035; vWFA.
 DR SMART; SM00327; vWA; 1.

DR PROSITE: PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE: PS0234; VWFA; 1.
SQ SEQUENCE 913 AA: 100070 MW: A7FA2F9E1089806D CRC64;

alignment_scores:
Quality: 319.00 Length: 76
Ratio: 4.493 Gaps: 0
Percent Similarity: 93.421 Percent Identity: 76.316

alignment_block:
US-09-049-696-10 x 088826 ..
Align seg 1/1 to: 088826 from: 1 to: 913

1 GGCACAGTGCCTGGACACACCGTGGAAAGGACACTTTGTTCTTAT 50
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
505 GlySerValIleValAspSerSerValGlyLysAspThrLeuPheLeuI 521
51 CACTTGGACAAACGAGCTCCCAATCTCTCTGGATCCAGTGGAC 100
||||:||||:||||:||||:||||:||||:||||:||||:||||:
521 eThrTrpThrHisProProThrIlePheIleTrpAspProSerGlyV 538
101 AGAAGCAAGTGGCTTTGTAGTGACAAAACACCAAAATGGCTACTC 150
||||:||||:||||:||||:||||:||||:||||:||||:||||:
538 alGluGlnAsnGlyPheIleLeuAspThrThrLysValAlaIleLeu 554
151 CAATCCAGGACATGCTAAGTTGGCAGCTTGGAAATACAGTCTGCAAGC 200
||||:||||:||||:||||:||||:||||:||||:||||:||||:
555 GlnValProGlyThrAlaLysValGlyPheTrpLysTyrSerIleGlnAl 571
201 AAGCTCACAACCTTGACCTGACTGTC 228
||||:||||:||||:||||:||||:||||:||||:||||:||||:
571 aSerSerGlnThrLeuThrLeuThrVal 580

seq_name: sp_rodent:Q9D726

seq_documentation_block:
ID Q9D726 PRELIMINARY; PRT; 913 AA.
AC Q9D726;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CHLORIDE CHANNEL CALCIUM ACTIVATED 3.
GN CLCA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=STOMACH;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saico T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saico R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection";
RL Nature 409:685-690(2001).

DR EMBL; AK008659; BAB25815.1; ..
DR MGD; MGI:1346342; Clca3.
DR InterPro; IPR00131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
SQ SEQUENCE 913 AA: 100111 MW: 491E584B06D9A89 CRC64;

alignment_scores:
Quality: 319.00 Length: 76
Ratio: 4.493 Gaps: 0
Percent Similarity: 93.421 Percent Identity: 76.316

alignment_block:
US-09-049-696-10 x Q9D726 ..
Align seg 1/1 to: Q9D726 from: 1 to: 913

1 GGCACAGTGCCTGGACACACCGTGGAAAGGACACTTTGTTCTTAT 50
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
505 GlySerValIleValAspSerSerValGlyLysAspThrLeuPheLeuI 521
51 CACTTGGACAAACGAGCTCCCAATCTCTCTGGATCCAGTGGAC 100
||||:||||:||||:||||:||||:||||:||||:||||:||||:
521 eThrTrpThrHisProProThrIlePheIleTrpAspProSerGlyV 538
101 AGAAGCAAGTGGCTTTGTAGTGACAAAACACCAAAATGGCTACTC 150
||||:||||:||||:||||:||||:||||:||||:||||:||||:
538 alGluGlnAsnGlyPheIleLeuAspThrThrLysValAlaIleLeu 554
151 CAATCCAGGACATGCTAAGTTGGCAGCTTGGAAATACAGTCTGCAAGC 200
||||:||||:||||:||||:||||:||||:||||:||||:||||:
555 GlnValProGlyThrAlaLysValGlyPheTrpLysTyrSerIleGlnAl 571
201 AAGCTCACAACCTTGACCTGACTGTC 228
||||:||||:||||:||||:||||:||||:||||:||||:||||:
571 aSerSerGlnThrLeuThrLeuThrVal 580

seq_name: sp_human:Q9NXP1

seq_documentation_block:
ID Q9NXP1 PRELIMINARY; PRT; 469 AA.
AC Q9NXP1;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE CDNA FLJ20131 FIS, CLONE COL06357.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project";
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK00138; BAA0969.1; ..
SQ SEQUENCE 469 AA; 50694 MW; 6765986C886B3E56 CRC64;

alignment_scores:
Quality: 263.00 Length: 77
Ratio: 3.925 Gaps: 1
Percent Similarity: 87.013 Percent Identity: 66.234

alignment_block:
US-09-049-696-10 x Q9NXP1

Tue Apr 2 09:39:37 2002

us-09-049-696-10.rspt

Align seg 1/1 to: Q9UNP1 from: 1 to: 469

4 ACAGTGATCGTGGACAGCACCGTGGGAAGGACACTTGTGTTCTTATCAC 53
 |||||.....
 58 ThrValIleIleAspSerThrValGlyLysAspThrPhePheLeuIleTh 74
 |||||.....
 54 CTGGACACGACGCTCCCAAAATCTTCTCTGGATCCCGAGTGGACAGA 103
 |||||.....
 74 rTrpAsnSerLeuProProSerIleSerLeuTrpAspProSerGlyThrI 91
 |||||.....
 104 AGCAAGTGGCTTTGTAGTGGACAAACACCAAAATGGCTACTCCAA 153
 ::::.....
 91 leMetGluAsnPhetrValAspAlaThrSerLysMetAlaThrLeuSer 107
 ::::.....
 154 ATCCAGGACATGCTAAGGTGGCTGGGAATACACTTGGCAAGCAAG 203
 |||||.....
 108 IleProGlyThrAlaLysValGlyThrTrpAlaTyAsnLeuGlnAlaLy 124
 ::::.....
 204 CTCAC.....CAACCTTGCACCTGACTGTC 228
 ::::.....
 124 sAlaAsnProGluThrLeuThrIleThrVal 134

seq_name: sp_human:Q9UNF7

seq_documentation_block:

ID Q9UNF7 PRELIMINARY; PRT; 917 AA.

AC Q9UNF7;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 2.
 GN CACCC2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COLON;
 RX MEDLINE=99364503; PubMed=10437792;
 RA Agnel M., Vernat T., Culouscou J.M.;
 RT "Identification of three novel members of the calcium-dependent
 RT chloride channel (CaCC) family predominantly expressed in the
 RT digestive tract and trachea.";
 RL FEBS Lett. 455:295-301(1999).
 DR EMBL; AF127035; AAD48398.1;
 DR InterPro; IPR002035; VWFA.
 DR Pfam; PF00092; VWFA; 1.
 DR PROSITE; PS50234; VWFA; 1.
 DR SMART; SM00327; VWFA; 1.
 SQ SEQUENCE 917 AA; 101153 MW; EA01090D781BEB95 CRC64;

alignment_scores:
 Quality: 263.00 Length: 77
 Ratio: 3.925 Gaps: 1
 Percent Similarity: 87.013 Percent Identity: 66.234

alignment_block:

US-09-049-696-10 x Q9UNF7

Align seg 1/1 to: Q9UNF7 from: 1 to: 917

4 ACAGTGATCGTGGACAGCACCGTGGGAAGGACACTTGTGTTCTTATCAC 53
 |||||.....
 506 ThrValIleIleAspSerThrValGlyLysAspThrPhePheLeuIleTh 522
 |||||.....
 54 CTGGACACGACGCTCCCAAAATCTTCTCTGGATCCCGAGTGGACAGA 103
 |||||.....
 522 rTrpAsnSerLeuProProSerIleSerLeuTrpAspProSerGlyThrI 539
 |||||.....
 104 AGCAAGTGGCTTTGTAGTGGACAAACACCAAAATGGCTACTCCAA 153
 ::::.....

seq_name: sp_mammal:O18742

seq_documentation_block:

ID O18742 PRELIMINARY; PRT; 794 AA.

539 leMetGluAsnPhetrValAspAlaThrSerLysMetAlaTyLeuSer 555
 ::::.....
 154 ATCCAGGACATGCTAAGGTGGCAATACACTTGGCAAGCAAG 203
 |||||.....
 556 IleProGlyThrAlaLysValGlyThrTrpAlaTyAsnLeuGlnAlaLy 572
 ::::.....
 204 CTCAC.....CAACCTTGCACCTGACTGTC 228
 ::::.....
 572 sAlaAsnProGluThrLeuThrIleThrVal 582

seq_name: sp_mammal:Q9TUB5

seq_documentation_block:

ID Q9TUB5 PRELIMINARY; PRT; 917 AA.

AC Q9TUB5;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE EPI-THELIAL CHLORIDE CHANNEL PROTEIN.
 GN AECC.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=ILEAL MUCOSA;
 RA Gaspar K.J., Gabriel S.E., Racette K.J., Forsyth G.W.;
 RT "A CDNA involved in porcine exocrine chloride conductance.";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF095584; AAF00077.1;
 DR InterPro; IPR000131; ATPase_gamma.
 DR InterPro; IPR002035; VWFA.
 DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
 DR PROSITE; PS50234; VWFA; 1.
 DR SMART; SM00327; VWFA; 1.
 SQ SEQUENCE 917 AA; 100735 MW; 649F41944F7EAD19 CRC64;

alignment_scores:
 Quality: 244.00 Length: 78
 Ratio: 3.812 Gaps: 1
 Percent Similarity: 82.051 Percent Identity: 66.667

alignment_block:

US-09-049-696-10 x Q9TUB5

Align seg 1/1 to: Q9TUB5 from: 1 to: 917

1 GCACAGTATCGTGGACAGCACCGTGGGAAGGACACTTGTGTTCTTAT 50
 |||||.....
 504 GlyThrValValAspSerThrValGlyLysAspThrLeuPheLeuI 520
 ::::.....
 51 CACCTGGACACCGCAG.....CCTCCCAAAATCTTCTCTGGATCCCA 94
 |||||.....
 520 eThrLeuGluArgLysPheLeuSerProIleProPheGlyValProS 537
 ::::.....
 95 GTGGACAGAACGAGTGGCTTTGTAGTGGACAAACACCAAAATGCC 144
 |||||.....
 537 erGlyArgSerGlnAspSerPheLeuValGlyLysHisAsnLysMetAla 553
 ::::.....
 145 TACCTCCAAATCCAGGCATTCCTAAGTTGGCACTTGGAAATACAGTCT 194
 |||||.....
 554 TyrPheGlnValProGlyThrAlaLysValGlyMetTrpLysTyrSerLe 570
 ::::.....
 195 GCAAGCAAGCTCACAAACCTTGACCTGACTGTC 228
 ::::.....
 570 uGlnAlaSerSerGlnThrLeuThrLeuThrVal 581

seq_name: sp_mammal:O18742

seq_documentation_block:

ID O18742 PRELIMINARY; PRT; 794 AA.

```

DR EMBL: AF001263; AAB86531.1; -.
DR InterPro: IPR002035; vWFA.
DR PROSITE; PS0234; vWFA; 1.
DR SMART: SM00327; vWA; 1.
DR SEQUENCE 820 AA; 91464 MW; 4E01330AEF5A1494 CRC64;

alignment_scores:
  Quality: 219.00      Length: 82
  Ratio: 3.476        Gaps: 3
  Percent Similarity: 76.829  Percent Identity: 59.756

alignment_block:
  US-09-049-696-10 x O18743  ..

  Align seg 1/1 to: O18743 from: 1 to: 820

      1  GCACAGTATCTGGACAGCACCTGGGAAGACACTTTGTTCTTAT 50
      506 GlyThrValProValAspSerThrValGlyAsnAspThrPhePheValVa 522
      51  CACCTGGACAAACACGCCCTCCCAATCCTTCTCTGGGACTCCAGCTGGAC 100
      522 lThrTrpThrIleGlnLysProGluIleValLeuGlnAspProLysGlyL 539
      101 AG.....AAGCAAGCTGCTTTGTAGTGGACAAA...AACACCAAAATG 141
      539 yslsytYrLysThrSerAspPheLysGluAspLysLeuAsnIleArgSer 555
      142 GCCTACTCTCAAAATCCAGGCAATCTCTAAGTTGGCTGGAATACAG 191
      556 AlaArgLeuGlnIleProGlyIleAlaGluThrGlyThrTrpThrYrSe 572
      192 TCTG .....CAAGCAAGCTCACAAACCTTGACCTGACTGCTC 228
      572 rLeuAsnAsnHisAlaSerSerGlnMetLeuThrValThrVal 587

seq_name: sp_mammal:O18741

seq_documentation_block:
  ID O18741 PRELIMINARY; PRT; 905 AA.
  AC O18741;
  DT 01-JAN-1998 (TReMBLrel. 05, Created)
  DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
  DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
  DE LU-ECAM-1.
  OS Bos taurus (Bovine).
  OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
  OC Bovidae; Bovinae; Bos.
  OC NCBI_TaxID=9913;
  RN [1]
  RP SEQUENCE FROM N.A.
  RC TISSUE=LUNG;
  RA Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
  RA Goodwin A., Pauli B.U.;
  RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
  EMBL; AF001261; AAB86529.1; -.
  DR InterPro: IPR002035; vWFA.
  DR InterPro: IPR003961; FN_III.
  DR Pfam; PF00041; fn3; 1.
  DR PROSITE; PS0234; vWFA; 1.
  DR SMART; SM00327; vWA; 1.
  DR SEQUENCE 905 AA; 101005 MW; D4CA6E7919115E88 CRC64;

alignment_scores:
  Quality: 219.00      Length: 82
  Ratio: 3.476        Gaps: 3
  Percent Similarity: 76.829  Percent Identity: 59.756

alignment_block:
  US-09-049-696-10 x O18741  ..

```

us-09-049-696-10.rspt

Tue Apr 2 09:39:37 2002

```
Align seg 1/1 to: 018741 from: 1 to: 905
1 GGCACAGTATCGTGGGACAGCACCGTGGGAAGGACACTTGTCTTAT 50
|||||
506 GlyThrValProValAspSerThrValGlyAsnAspThrPhePheVal 522
|||||
51 CACCTGGGACAGCGCTCCCAATCTTCTCTGGGATCCAGTGGAC 100
|||||
522 IthrTrpThrLeuGlnLysProGluLeuValLeuGlnAspProLysGlyL 539
|||||
101 AG.....AAGCAAGTGGCTTTGTAGTGGACAAA...AACACCAAAATG 141
|||
539 yslYsYrLysThrSerAspPheLysGluAspLysLeuAsnIleArgSer 555
|||||
142 GCCTACCTCAATCCAGGATCTTAAAGTTGGCACTTGGAAATACAG 191
|||||
556 AlaArgLeuGlnIleProGlyLeuAlaGluThrGlyThrTrpTyrSe 572
|||||
192 TCTG.....CAAGCAAGCTCACAACCTTGACCTGACTGTC 228
|||||
572 rLeuLeuAsnAsnHisAlaSerSerGlnMetLeuThrValThrVal 587
|||||
seq_name: sp_rodent:Q9R070

seq_documentation_block:
ID Q9R070 PRELIMINARY; PRT; 902 AA.
AC Q9R070;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CA(2+)-SENSITIVE CHLORIDE CHANNEL 2.
GN CLCA2 OR CACC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57/BL6;
RA Elble R.C., Pauli B.;
RT "A murine endothelial homolog of Lu-ECAM-1 associated with lung
RT melanoma metastasis.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF115852; AAG47626.1;
DR InterPro; IPR003961; FN.III.
DR SMART; SM0060; FN3; 1.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS50234; VWFA; 1.
DR SEQUENCE 902 AA; 99866 MW; 0FB746CF5FB3D07F CRC64;
SQ

alignment_scores:
Quality: 201.00 Length: 79
Ratio: 3.242 Gaps: 3
Percent Similarity: 78.481 Percent Identity: 54.430

alignment_block:
US-09-049-696-10 x Q9R070
Align seg 1/1 to: Q9R070 from: 1 to: 902
4 ACAGTGATCGTGGGACAGCACCGTGGGAAGGACACTTGTCTTATCAC 53
|||||
506 ThrValLeuValAspSerThrValGlyAsnAspThrPhePheValIleTh 522
|||||
54 CTGGACAAGCGACCTCCCAATCTTCTCTGGGATCCAGTGGACAGA 103
|||||
522 rTrpThrValGlnLysProGluLeuIleLeuGlnAspProLysGlyLysL 539
|||||
104 AG.....CAAGTGGCTTTGTAGTGGACAAA...AACACCAAAATGCC 141
|||
539 yslYrLysThrSerAspPheLysGluAspLysLeuAsnIleArgSerAla 555
|||||
145 TACCTCCAAATCCAGGATCTTAAAGTTGGCACTTGGAAATACAGTCT 194
|||||
556 ArgLeuGlnIleProGlyThrAlaGluThrGlyThrTrpTyrSerIl 572
|||||
195 GCAAGCAAGC...TCACAAACCTTGACCTGACTGTC 228
|||||
572 eThrGlyThrLysSerGlnLeuIleThrMetThrVal 584
|||||
seq_name: sp_rodent:088860

seq_documentation_block:
ID Q9EQR4 PRELIMINARY; PRT; 902 AA.
AC Q9EQR4;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE ENDOTHELIAL CHLORIDE CHANNEL.
GN CLCA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57/BL6;
RA Elble R.C., Pauli B.;
RT "A murine endothelial homolog of Lu-ECAM-1 associated with lung
RT melanoma metastasis.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF115852; AAG47626.1;
DR InterPro; IPR003961; FN.III.
DR SMART; SM0060; FN3; 1.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS50234; VWFA; 1.
DR SEQUENCE 902 AA; 99855 MW; F2BA57DAA5D4AAAD CRC64;
SQ

alignment_scores:
Quality: 196.00 Length: 79
Ratio: 3.213 Gaps: 3
Percent Similarity: 77.215 Percent Identity: 54.430

alignment_block:
US-09-049-696-10 x Q9EQR4
Align seg 1/1 to: Q9EQR4 from: 1 to: 902
4 ACAGTGATCGTGGGACAGCACCGTGGGAAGGACACTTGTCTTATCAC 53
|||||
506 ThrValProValAspSerThrValGlyAsnAspThrPhePheValIleTh 522
|||||
54 CTGGACAAGCGACCTCCCAATCTTCTCTGGGATCCAGTGGACAGA 103
|||||
522 rTrpThrValGlnLysProGluLeuIleLeuGlnAspProLysGlyLysL 539
|||||
104 AG.....CAAGTGGCTTTGTAGTGGACAAA...AACACCAAAATGCC 144
|||
539 yslYrLysThrSerAspPheLysGluAspLysLeuAsnIleArgSerAla 555
|||||
145 TACCTCCAAATCCAGGATCTTAAAGTTGGCACTTGGAAATACAGTCT 194
|||||
556 ArgLeuGlnIleProGlyThrAlaGluThrGlyThrTrpTyrSerIl 572
|||||
195 GCAAGCAAGC...TCACAAACCTTGACCTGACTGTC 228
|||||
572 eThrGlyThrLysSerGlnLeuIleThrMetThrVal 584
|||||
seq_name: sp_rodent:088860

seq_documentation_block:
ID Q9R070 PRELIMINARY; PRT; 902 AA.
AC Q9R070;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CA(2+)-SENSITIVE CHLORIDE CHANNEL 2.
GN CLCA2 OR CACC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57/BL6;
RA Elble R.C., Pauli B.;
RT "A murine endothelial homolog of Lu-ECAM-1 associated with lung
RT melanoma metastasis.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF108501; AAF12731.1;
DR MGD; MGI:1931471; Clca2.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS50234; VWFA; 1.
DR SEQUENCE 902 AA; 99866 MW; 0FB746CF5FB3D07F CRC64;
SQ

alignment_scores:
Quality: 201.00 Length: 79
Ratio: 3.242 Gaps: 3
Percent Similarity: 78.481 Percent Identity: 54.430

alignment_block:
US-09-049-696-10 x Q9R070
Align seg 1/1 to: Q9R070 from: 1 to: 902
4 ACAGTGATCGTGGGACAGCACCGTGGGAAGGACACTTGTCTTATCAC 53
|||||
506 ThrValLeuValAspSerThrValGlyAsnAspThrPhePheValIleTh 522
|||||
54 CTGGACAAGCGACCTCCCAATCTTCTCTGGGATCCAGTGGACAGA 103
|||||
522 rTrpThrValGlnLysProGluLeuIleLeuGlnAspProLysGlyLysL 539
|||||
104 AG.....CAAGTGGCTTTGTAGTGGACAAA...AACACCAAAATGCC 144
|||
539 yslYrLysThrSerAspPheLysGluAspLysLeuAsnIleArgSerAla 555
|||||
145 TACCTCCAAATCCAGGATCTTAAAGTTGGCACTTGGAAATACAGTCT 194
|||||
556 ArgLeuGlnIleProGlyThrAlaGluThrGlyThrTrpTyrSerIl 572
|||||
195 GCAAGCAAGC...TCACAAACCTTGACCTGACTGTC 228
|||||
572 eThrGlyThrLysSerGlnLeuIleThrMetThrVal 584
|||||
seq_name: sp_rodent:088860

seq_documentation_block:
ID Q9R070 PRELIMINARY; PRT; 902 AA.
AC Q9R070;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CA(2+)-SENSITIVE CHLORIDE CHANNEL 2.
GN CLCA2 OR CACC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57/BL6;
RA Elble R.C., Pauli B.;
RT "A murine endothelial homolog of Lu-ECAM-1 associated with lung
RT melanoma metastasis.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF108501; AAF12731.1;
DR MGD; MGI:1931471; Clca2.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS50234; VWFA; 1.
DR SEQUENCE 902 AA; 99866 MW; 0FB746CF5FB3D07F CRC64;
SQ
```

```

ID O88860 PRELIMINARY; PRT; 901 AA.
AC O88860;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CHLORIDE CHANNEL CACC.
GN CLCAL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Romio L., Musante L., Cinti R., Moran O., Seri M., Galletta L.J.V.;
RT "Characterization of a murine gene homologous to the bovine Cacc
RL Chloride channel.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF052746; AAC35003.1; -.
DR MGD; MGI:1316732; Clcal.
DR InterPro; IPR002035; vWFA.
DR SMART; SM00327; vWA.1.
DR PROSITE; PS0234; vWFA.1.
SQ SEQUENCE 901 AA; 100038 MW; 13DF068BF3337AAB CRC64;

alignment_scores:
  Quality: 193.00 Length: 80
  Ratio: 3.164 Gaps: 3
  Percent Similarity: 76.250 Percent Identity: 52.500

alignment_block:
US-09-049-696-10 x O88860 ..
  Align seg 1/1 to: O88860 from: 1 to: 901

1 GCACAGTGTGTCGGACAGCAGCCGTGGGAAAGACACTTTGTTCTTAT 50
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
505 GlyThrValProLeuAspSerThrValGlyAsnAspThrPheValI 521

51 CACCTGGACAAGCAGCCCTCCCAATCCTCTCTGGGATCCAGTGGAC 100
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
521 eHrTrpMetValLysProGluIleLeuGlnAspProLysGlyL 538

101 AGAAG.....CAAGTGGCTTTGTAGTGACAAA...AACACCAAAATG 141
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
538 ySLyTyrThrThrSerAspPheGlnAspLysLeuAsnIleArgSer 554

142 GCCTACCTCCAAATCCAGGCATTGCTAAGTTGGCATTGGAAATACAG 191
||| ||||| :||| :||| :||| :||| :||| :||| :||| :|||
555 AlaArgLeuGlnIleProGlyThrAlaGluThrGlyThrTrpThrTy 571

192 TGTGCAAGCAAGC...TCACAAACCTTGACCTGACTGTC 228
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
571 rIleThrGlyThrLysSerGlnLeuIleThrMetThrVal 584

seq_name: sp_rodent:O90X15

seq_documentation_block:
ID O90X15 PRELIMINARY; PRT; 902 AA.
AC O90X15;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIUM-SENSITIVE CHLORIDE CONDUCTANCE PROTEIN-1.
GN CLCAL OR MCICAL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-WHOLE LUNG.
RX MEDLINE=99041980; PubMed=9822685;

```

```

PA Gandhi R., Elble R.C., Gruber A.D., Schreur K.D., Ji H.-L.,
RA Fuller C.W., Pauli B.O.;
RT "Molecular and functional characterization of a calcium-sensitive
RL chloride channel from mouse lung.";
RL J. Biol. Chem. 273:32096-32101(1998).
DR EMBL; AF047838; AAC79982.1; -.
DR MGD; MGI:1316732; Clcal.
DR InterPro; IPR002035; vWFA.
DR SMART; SM00327; vWA.1.
DR PROSITE; PS0234; vWFA.1.
SQ SEQUENCE 902 AA; 100107 MW; D9A915E698EA2D85 CRC64;

alignment_scores:
  Quality: 193.00 Length: 80
  Ratio: 3.164 Gaps: 3
  Percent Similarity: 76.250 Percent Identity: 52.500

alignment_block:
US-09-049-696-10 x O90X15 ..
  Align seg 1/1 to: O90X15 from: 1 to: 902

1 GCACAGTGTGTCGGACAGCAGCCGTGGGAAAGACACTTTGTTCTTAT 50
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
505 GlyThrValProLeuAspSerThrValGlyAsnAspThrPheValI 521

51 CACCTGGACAAGCAGCCCTCCCAATCCTCTCTGGGATCCAGTGGAC 100
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
521 eHrTrpMetValLysProGluIleLeuGlnAspProLysGlyL 538

101 AGAAG.....CAAGTGGCTTTGTAGTGACAAA...AACACCAAAATG 141
:|||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
538 ySLyTyrThrThrSerAspPheGlnAspLysLeuAsnIleArgSer 554

142 GCCTACCTCCAAATCCAGGCATTGCTAAGTTGGCATTGGAAATACAG 191
||| ||||| :||| :||| :||| :||| :||| :||| :||| :|||
555 AlaArgLeuGlnIleProGlyThrAlaGluThrGlyThrTrpThrTy 571

192 TGTGCAAGCAAGC...TCACAAACCTTGACCTGACTGTC 228
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
571 rIleThrGlyThrLysSerGlnLeuIleThrMetThrVal 584

```

This Page Blank (uspto)

OM of: US-09-049-696-10 to: SwissProt_39:* out_format : pfs
Date: Mar 30, 2002 2:52 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+nt2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09049696/runat_28032002.145238_2085/app_query.fasta.1.12579
-DB=SwissProt_39 -OFMT=fastan -SUFFIX=rs -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPT=0.000 -LOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPEXT=7.000 -FGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGNL_1_165 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-10
Query length: 229
Database: SwissProt_39.*
Database sequences: 100059
Database length: 3664827
Search time (sec): 306.030000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
SwissProt_39:ECIC_BOVIN +	205.00	406.83	6.1e-16	903	1	P54281 bos taurus (bovine).
SwissProt_39:YMFQ_ECOLI +	70.00	134.93	2.91	263	1	P75981 escherichia coli. hyp
SwissProt_39:LAG3_MOUSE +	69.50	128.03	3.56	521	1	Q61790 mus musculus (mouse).
SwissProt_39:HBA_AMBE -	67.00	133.92	6.13	142	1	P2015 ambystoma mexicanum (a
SwissProt_39:YMFQ_ECOLI -	65.00	127.07	10.81	194	1	P75982 escherichia coli. hyp
SwissProt_39:VLLI_COV +	65.00	118.92	11.86	503	1	Q89828 canine oral papillomavi
SwissProt_39:NCAL_XENLA +	65.00	112.29	12.79	1092	1	P36335 xenopus laevis (afri
SwissProt_39:BP28_DROME +	64.00	104.62	17.82	2096	1	Q9vm75 drosophila melanogast
SwissProt_39:RPOB_RICN -	63.50	107.19	19.56	1373	1	Q9rh41 rickettsia conorii.
SwissProt_39:RPOB_RICMA -	63.50	107.19	19.56	1373	1	Q9rh43 rickettsia massillae
SwissProt_39:RPOB_RICR -	63.50	107.18	19.56	1374	1	Q52271 rickettsia prowazeki
SwissProt_39:RPOB_RICR -	63.50	107.17	19.57	1376	1	P77941 rickettsia typhi. dr
SwissProt_39:HEX_ADEL16 +	63.00	109.38	21.56	940	1	P36854 human adenovirus type
SwissProt_39:NCAL_XENLA +	63.00	108.13	21.87	1088	1	P16170 xenopus laevis (afri
SwissProt_39:ENV_BLBV2 -	62.50	113.48	23.26	515	1	P25506 bovine leukemia virus
SwissProt_39:DAG1_RABIT +	62.50	108.76	24.54	895	1	Q28685 oryctolagus cuniculus
SwissProt_39:P2X2_CAVPO +	61.50	112.10	30.17	474	1	O70397 cavia porcellus (guin
SwissProt_39:CY41_TRYBB +	61.50	103.86	33.13	1242	1	Q99279 trypanosoma brucei b
SwissProt_39:GIL2_ENTHI +	61.00	115.32	32.87	288	1	Q30377 entamoeba histolytica
SwissProt_39:DHAL_ASPNG -	61.00	110.65	34.66	497	1	P41751 aspergillus niger. al
SwissProt_39:DNBI_EBV +	61.00	103.64	37.54	1128	1	P33227 Epstein-Barr virus (st
SwissProt_39:CAML_RAT +	61.00	102.70	37.94	1259	1	Q05695 rattus norvegicus (rat
SwissProt_39:NGCA_CHICK +	61.00	102.65	37.96	1266	1	Q03696 gallus gallus (chick
SwissProt_39:GTFD_STRMU +	61.00	101.42	38.50	1462	1	P49331 streptococcus mutans
SwissProt_39:LAR_DROME +	61.00	98.62	39.74	2029	1	P16621 drosophila melanogast
SwissProt_39:TEGU_HSV7J -	61.00	98.49	39.80	2059	1	P52362 human herpesvirus (t
SwissProt_39:MEI1_HUMAN +	60.50	111.68	38.72	390	1	Q00470 homo sapiens (human)
SwissProt_39:MEI1_MOUSE +	60.50	111.68	38.72	390	1	Q05954 mus musculus (mouse).
SwissProt_39:TGLH_TACTR +	60.50	105.93	41.33	764	1	Q35818 tachyplesus tridentatu
SwissProt_39:HEX_ADE04 +	60.00	109.47	44.87	447	1	P05187 human adenovirus type
SwissProt_39:CORO_SCHPO +	60.00	106.93	46.18	601	1	Q13923 schizosaccharomyces p
SwissProt_39:DAG1_BOVIN +	60.00	103.53	48.00	895	1	Q18738 bos taurus (bovine).
SwissProt_39:CAML_MOUSE +	60.00	100.60	49.62	1260	1	P11627 mus musculus (mouse)
SwissProt_39:HV20_MOUSE +	59.50	119.53	45.23	122	1	P01789 mus musculus (mouse)
SwissProt_39:RACD_DICDI +	59.50	113.25	48.57	254	1	P34150 dictyostelium discoide
SwissProt_39:CD68_HUMAN +	59.50	110.42	50.16	354	1	P34810 homo sapiens (human).
SwissProt_39:ENV_BLV -	59.50	107.21	52.02	515	1	P51519 bovine leukemia virus (b
SwissProt_39:ENV_BLBV5 -	59.50	107.21	52.02	515	1	P25507 bovine leukemia virus
SwissProt_39:DNK_HUMAN -	59.50	105.36	53.12	639	1	Q09013 homo sapiens (human).
SwissProt_39:YANK_RHISN +	59.50	105.06	53.31	662	1	P55583 rhizobium sp. (strain

alignment_scores:
Quality: 205.00 Length: 82

SwissProt_39:CCAD_MESAU + 59.50 97.46 58.12 1610 ! Q99244 mesocricetus aura
SwissProt_39:CCAD_HUMAN + 59.50 94.94 59.80 2161 ! Q01668 homo sapiens (hum
SwissProt_39:CCAD_RAT + 59.50 94.77 59.92 2203 ! P27732 rattus norvegicus (r
SwissProt_39:SACS_MOUSE + 59.50 90.04 63.23 3830 ! O91lc8 mus musculus (mou
SwissProt_39:YFSS_RHILP + 59.00 120.62 50.48 95 ! P10497 rhizobium leguminos
seq_name: SwissProt_39:ECIC_BOVIN
seq_documentation_block:
ID ECIC_BOVIN STANDARD; PRT; 903 AA.
AC P54281;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN (CALCIUM-ACTIVATED CHLORIDE
CHANNEL).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Trachea;
RX MEDLINE=96125078; PubMed=8537359;
RA Cunningham S.A., Awayda M.S., Bublief J.K., Ismailov I.I.,
RA Arrate M.P., Berdiev B.K., Benos D.J., Fuller C.M.;
RT "Cloning of an epithelial chloride channel from bovine trachea.";
RL J. Biol. Chem. 270:31016-31026(1995).
CC !- FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE
SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME;
MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND
TRANSEPITHELIAL TRANSPORT.
CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC !- TISSUE SPECIFICITY: TRACHEA.
CC !- PTM: PHOSPHORYLATED BY CAM-KINASE II.
CC !- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U36445; AAC48511.1; -
DR InterPro: IPR002035; VWFA.
DR SMART: SM00327; VWFA; 1.
DR PROSITE: PS50234; VWFA; 1.
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
Calcium channel; Phosphorylation; Glycoprotein.
FT TRANSMEM 7 27
FT TRANSMEM 331 351
FT TRANSMEM 617 637
FT TRANSMEM 883 903
FT DOMAIN 308 476
FT CARBOHYD 75 75
FT CARBOHYD 278 278
FT CARBOHYD 360 360
FT CARBOHYD 372 372
FT CARBOHYD 504 504
FT CARBOHYD 515 515
FT CARBOHYD 688 688
FT CARBOHYD 811 811
FT CARBOHYD 816 816
FT CARBOHYD 842 842
FT CARBOHYD 857 857
SQ SEQUENCE 903 AA; 100305 MW; 9742AB1590908AEC CRC64;

72 eSerAspSerSerAsnIleAlaAsnLeuGlu.....P 83
128 TTGTCCACTAC..... 118
83 heValLlyTyrGluPheAspThrProLysTyrAspValGluGluCysSer 99
117AAAGCCACTTGTCTTCTGTC 98

100 GlnArgSerLeuSerTyrAlaAlaProLeuLysValThrLeuArgLeuSe 116
97 CACTGGGATCCAGAGAGATTGGGGAGCGCTGCTGTCAGGTGATA 48
116 rleTtpAspIleAspGluAspThrGlyThrArgGluIleLysGlyLeL 133
47 AGAACAAAGTGC 34
133 ysGluGlnGluVal 137

seq_name: SwissProt_39:RPOB_RICMA

seq_documentation_block:
ID RPOB_RICMA STANDARD; PRT; 1373 AA.
AC Q9RH43;
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6) (TRANSCRIPTASE
DE BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT).
GN RPOB.
OS Rickettsia massiliae.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35791;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MTUI;
RX MEDLINE=99437773; PubMed=10508014;
RA Drancourt M., Raoult D.;
RT "Characterization of mutations in the rpoB gene in naturally
RT rifampin-resistant Rickettsia species.",
RL Antimicrob. Agents Chemother. 43:2400-2403(1999).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE +
CC RNA(N).
CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN.
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF076433; AAF22431.1; -.
DR InterPro; IPR001572; RNA_pol_B.
DR Pfam; PF00562; RNA_pol_B; 1.
DR PROSITE; PS01166; RNA_POL_BETA; 1.
KW Transferase; DNA-directed RNA polymerase; Transcription.
SQ SEQUENCE 1373 AA; 154278 MW; C530FE629F99CACE CRC64;

alignment_scores:
Quality: 63.50 Length: 88
Ratio: 1.587 Gaps: 2
Percent Similarity: 45.455 Percent Identity: 26.136

alignment_block:

US-09-049-696-10/rev x RPOB_RICMA ..

Align seg 1/1 to: RPOB_RICMA from: 1 to: 1373

228 GACAGTCAGGTCAGGTTGTGAGCTGCTGTCAGACTGATTTCACAG 179
::: ::::::::::::::::::::|:::|:::|:::|:::|:::|:::|:::|
56 GluArgLysAsnLysGlyLeuGlnSerIleLeuAsnSerIlePheProL 72

178 TGCCCAACCTTAGCAATGCGCTGGGATTGGAGGTAGGCCATTTGGTGTTT 129
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
72 eSerAspSerSerAsnIleAlaAsnLeuGlu.....P 83
128 TTGTCACACTAC..... 118
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
83 heValLysTyrGluPheAspThrProLysTyrAspValGluGluCysSer 99
117 98
100 GlnArgSerLeuSerTyrAlaAlaProLeuLysValThrLeuArgLeuSe 116
97 CACTGGGATCCAGAGAGATTGGGGAGCGCTGCTGTCAGGTGATA 48
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
116 rleTtpAspIleAspGluAspThrGlyThrArgGluIleLysGlyLeL 133
47 AGAACAAAGTGC 34
133 ysGluGlnGluVal 137

seq_name: SwissProt_39:RPOB_RICPR

seq_documentation_block:
ID RPOB_RICPR STANDARD; PRT; 1374 AA.
AC O52271; Q9RH37;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6) (TRANSCRIPTASE
DE BETA CHAIN) (RNA POLYMERASE BETA SUBUNIT).
GN RPOB OR RP140.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MADRID E;
RX MEDLINE=98215179; PubMed=9555894;
FA Rachek L.I., Tucker A.M., Winkler H.H., Wood D.O.;
RT "Transformation of Rickettsia prowazekii to rifampin resistance.",
RL J. Bacteriol. 180:2118-2124(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sichertitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.",
RL Nature 396:133-140(1998).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=BREINL;
RX MEDLINE=99437773; PubMed=10508014;
RA Drancourt M., Raoult D.;
RT "Characterization of mutations in the rpoB gene in naturally
RT rifampin-resistant Rickettsia species.",
RL Antimicrob. Agents Chemother. 43:2400-2403(1999).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE +
CC RNA(N).
CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN.
CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

97 CACTGGGATCCAGACAGGATTGGGAGGCTGCTTCTCCAGGTGATA 48
 116 rleTrpAspIleAspGluAspThrGlyThrArgGluIleLysGlyIleL 133
 47 AGAAACAAAGTGC 34
 133 ySgLuGluGluVal 137

seq_name: SwissProt_39:HEX_ADE16

seq_documentation_block:
 ID HEX_ADE16 STANDARD; PRT; 940 AA.

AC P36854;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HEXON PROTEIN (LATE PROTEIN 2).
 GN PII.

OS Human adenovirus type 16.
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 OX NCBI_TaxID=31544;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=ISOLATE CH.79;
 RX MEDLINE=95407102; PubMed=7676636;
 RA Pring-Akerblom P., Trijssenaar J., Adrian T.;
 RT "Sequence characterization and comparison of human adenovirus
 subgenus B and E hexons";
 RL Virology 212:232-236(1995).

CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE STRUCTURAL PROTEINS IN THE
 VIRAL COAT AND IS SYNTHESIZED DURING LATE INFECTION.

CC -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

CC EMBL: X74662; CAA52726.1; -
 DR PIR: S37216; S37216.

DR HSP: P03277; IDHX

DR InterPro: IPR000736; Adeno_hexon.
 DR Pfam: PF01065; Adeno_hexon; 1.

DR ProDom: PD002815; Adeno_hexon; 1.
 DR CoaT protein; Hexon protein; Late protein.

DR CoaT protein; Hexon protein; Late protein.
 SQ SEQUENCE 940 AA; 105501 MW; E5A3493E68236D15 CRC64;

alignment_scores:

Quality: 63.00 Length: 49
 Ratio: 2.172 Gaps: 4
 Percent Similarity: 59.184 Percent Identity: 40.816

alignment_block:

US-09-049-696-10 x HEX_ADE16 ..

Align seg 1/1 to: HEX_ADE16 from: 1 to: 940

13 GTGGACACACCGTGGGAAGACACTTGTTCATTATCACCCTGGACAC 62
 ::::::::::: ||| ::::::::::: |||
 167 ileAspSerThrSerGlyThrAspThrValIleTyAlaAspLysThrPh 183

63 GCAGCCT...CCCCAATC.....CTTCTCTGG...GATCCCACTGGA. 99
 ||||| ||||| ::::::::::: ||| ||| ::::::::::: |||
 183 eGlnProGluProGlnValGlyAsnAlaSerTrpValAspAlaAsnGlyT 200

100CAGAAGCAAGTGGCTTTGTAGTGGCAAAACACCAAAATG 141
 ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

200 hrGluGluLysTyrglyGlyArgAlaLeuLysAspThrThrLysMet 215

seq_name: SwissProt_39:NCAL_XENLA

seq_documentation_block:

ID NCAL_XENLA STANDARD; PRT; 1088 AA.
 AC P16170;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE NEURAL CELL ADHESION MOLECULE 1, LARGE ISOFORM PRECURSOR (N-CAM 180)
 [CONTAINS: N-CAM 140].

DR NCAM1.

GN NCAM1.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI_TaxID=8355;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=90098871; PubMed=2481269;

RA Krieg P.A., Sakaguchi D.S., Kintner C.R.;

RT "Primary structure and developmental expression of a large

cytoplasmic domain form of Xenopus laevis neural cell adhesion

molecule (NCAM).";

RL Nucleic Acids Res. 17:10321-10335(1989).

CC -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN

NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF

NEURITES, ETC.

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF

N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -!- TISSUE SPECIFICITY: EXPRESSED IN NEURON AND IN PRESUMPTIVE NEURAL

TISSUE.

CC -!- DEVELOPMENTAL STAGE: THE MRNA ENCODING THIS LD-NCAM IS THE MAJOR

TRANSCRIPT PRESENT IN BOTH MATERNAL RNA AND IN THE EMBRYO DURING

EARLY NEURAL DEVELOPMENT.

CC -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.

CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

CC EMBL: M25696; AAA49909.1; -

DR PIR: S09600; IJXLNL.

DR HSP: P56276; IYTK.

DR InterPro: IPR001777; FN_III.

DR InterPro: IPR003006; Ig_MHC.

DR InterPro: IPR003598; Ig_C2.

DR Pfam: PF00041; fn3; 2.

DR Pfam: PF00047; Ig; 5.

DR SMART: SM00060; FN3; 2.

DR SMART: SM00408; IGC2; 5.

DR Cell adhesion; Glycoprotein; Transmembrane; Repeat; Brain;

KW Immunoglobulin domain; Alternative splicing; Signal.

FT SIGNAL 1 19

FT CHAIN 20 1088

FT DOMAIN 20 705

FT TRANSMEM 706 723

FT DOMAIN 724 1088

FT DOMAIN 34 100

FT DOMAIN 129 193

FT DOMAIN 225 289

FT DOMAIN 316 386

FT DOMAIN 413 480

FT DOMAIN 512 589

FT DOMAIN 518 686

FT DOMAIN 149 153

FT DOMAIN 20 705

FT DOMAIN 724 1088

FT DOMAIN 34 100

FT DOMAIN 129 193

FT DOMAIN 225 289

FT DOMAIN 316 386

FT DOMAIN 413 480

FT DOMAIN 512 589

FT DOMAIN 518 686

FT DOMAIN 149 153

```

FT DOMAIN 158 162 HEPARIN-BINDING (POTENTIAL).
FT DISULFID 41 93 PROBABLE.
FT DISULFID 136 186 PROBABLE.
FT DISULFID 232 282 PROBABLE.
FT DISULFID 323 379 PROBABLE.
FT DISULFID 420 473 PROBABLE.
FT CARBOHYD 82 82 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 341 341 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 804 1049 MISSING (IN ISOFORM N-CAM 140).
SQ SEQUENCE 1088 AA; 117778 MW; 62738B5B03F3E83 CRC64;

```

```

alignment_scores:
  Quality: 63.00 Length: 60
  Ratio: 1.703 Gaps: 3
Percent Similarity: 61.667 Percent Identity: 33.333

```

alignment_block:

US-09-049-696-10 x NCAL_XENLA ..

Align seg 1/1 to: NCAL_XENLA from: 1 to: 1088

```

28 GGAAGGACACTTTGTTTTCATACCTGGACACGAGCCTCCCAAAAT 77
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
129 GlyGluAspAlaValIleCysAspValSerSerIleProSerIle 145

78 CCTTCTCTGGATCCCATGGA.....CAGNAGCAAGGTG 112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145 eilethTrpArgHisGlyLysAspValIlePheLysAspVala 162

113 GCTTTGATGGACAAAACACCAAAATGGCTACCTCCAAATCCCAAGC 162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
162 rgPheValValLeuAlaAsnAsn.....TyrLeuGlnIleArgGly 175

163 ATTCTCAAGTT.....GGCACTTGGA 186
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
176 IleLysLysThrAspGluGlyThrTyrArg 185

```

seq_name: SwissProt_39:ENV_BLVB2

seq_documentation_block:

```

ID ENV_BLVB2 STANDARD; PRT; 515 AA.
AC P25506;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ENV POLYPROTEIN PRECURSOR [CONTAINS: COAT GLYCOPROTEIN GP51; COAT
DE PROTEIN GP30].
GN ENV.
OS Bovine leukemia virus (Belgium isolate LB285) (BLV).
OC Viruses; Retroid viruses; Retroviridae; Deltaretrovirus.
OX NCBI_TaxID=11905;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90347810; PubMed=1696635;
RA Mamoun R., Morisson M., Rebeyrotte N., Busetta B., Couez D.,
RA Kettmann R., Hospital M., Guillemain B.;
RT "Sequence variability of bovine leukemia virus env gene and its
RT relevance to the structure and antigenicity of the glycoproteins.";
RL J. Virol. 64:4180-4188(1990).
CC -!- SIMILARITY: VERY STRONG, WITH BOVINE LEUKEMIA VIRUS ENV
CC POLYPROTEIN FROM OTHER ISOLATES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC -----
DR HMBL; M35240; AAA42791.1; ..
DR HMBL; P03385; 1MOF.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; ENV_polyprotein; 1.
KW Glycoprotein; Coat protein; Transmembrane; Polyprotein; Signal.
FT SIGNAL 1 33
FT CHAIN 34 301 COAT GLYCOPROTEIN GP51.
FT CHAIN 302 515 COAT PROTEIN GP30.
FT TRANSMEM 304 324 POTENTIAL.
FT TRANSMEM 436 456 POTENTIAL.
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 203 203 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 515 AA; 58580 MW; 84075568CBE9A5CA CRC64;

```

alignment_scores:

```

  Quality: 62.50 Length: 110
  Ratio: 1.453 Gaps: 6
Percent Similarity: 39.091 Percent Identity: 25.455

```

alignment_block:

US-09-049-696-10/rev x ENV_BLVB2 ..

Align seg 1/1 to: ENV_BLVB2 from: 1 to: 515

```

214 AGTTTGTGAGCTGCTGCAGACTGATTTCCAAGTCCCAACCTTAGCA 165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16 ArgTrpValSerLeuThrLeuThrLeuSerLeuGlnProIleG1 32

164 A.....TGCTGGGATTGGAGGTAGGCCATTTTGGTGT..... 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32 nThrTrpArgCysSerLeuSerLeuGlyAsnGlnGlnTrpMetThrThrt 49
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 ..... 131

49 yrAsnGlnGluAlaLysPheSerIleAlaIleAspGlnIleLeuGluAla 65

130 .....TTTGTCCACTACAAAGCCACCTTGCT..... 104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
66 HisAsnGlnSerProPheCysProArgSerProArgTyrThrLeuAspPh 82

103 .....TCTGTCCACTGGATCCAGACAGAGA..... 77
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
82 eValAsnGlyTyrProLysIleTyrTrpProProGlnGlyArgArga 99

76 ..TTTGGG...GAGGCTGGGTTCAGGAGTAAAGAAACAAAGTGCCT 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
99 rgPheGlyAlaArgAlaMetValThrTyrAspCysGluProArgCysPro 115

31 TTCCACGCGTGTGCTCCACGATCACTGT 4
::: ||||| |||
116 TyrVal.GlyAlaAspHisPheAspCys 124

```


OM of: US-09-049-696-10 to: PIR_68:* out_format : pfs
Date: Mar 30, 2002 2:26 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+ntp.model -DEV=xlp
-O=/cgen2_1/USPRO.spool/US09049696/runat_28032002_145238_2031/app_query.fasta_1.12579
-DB=PIR_68 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPEXT=0.050 -GAPOP=10.000 -XGAPEXT=0.500 -XGAPEXT=0.500 -XGAPEXT=0.500
-DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -OUTSCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -THRESH=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09049696_@CGN1_1_367 -NCPY=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-049-696-10
Query length: 229
Database: PIR_68*
Database sequences: 219241
Database length: 7617452
Search time (sec): 470.790000

score_list:	Strd Orig	ZScore	EScore	Len	Documentation
Sequence					
pir2:T06246	+ 319.00	629.51	4.9e-28	913	! gob-5 protein - mouse
pir2:T06242	+ 219.00	425.94	1.1e-16	905	! Lu-ECAM-1 protein - mouse
pir2:T45935	- 73.00	132.32	4.08	545	! probable galactose oxidase (EC
pir2:T64860	- 70.00	131.50	9.38	263	! hypothetical protein b1152 - Es
pir1:HAXM	- 67.00	129.87	21.43	142	! hemoglobin alpha chain - axolot
pir2:S78549	- 67.00	109.57	17.71	2321	! notch3 protein - human
pir2:PH4215	+ 66.00	124.43	26.95	227	! Ig epsilon chain C region form
pir4:A47271	+ 65.50	128.41	32.18	114	! nitrophenyl phosphonate-specifi
pir2:S20901	+ 65.50	98.71	24.34	6805	! titin - rabbit (fragment)
pir2:T64860	- 65.00	123.53	35.36	194	! hypothetical protein b1153 - Es
pir2:JE0100	+ 65.00	113.95	32.32	725	! neural cell adhesion molecule 2
pir1:JN0635	+ 65.00	110.98	31.43	1092	! neural cell adhesion molecule 1
pir1:J83844	- 64.50	87.69	25.25	26926	! titin, cardiac muscle [valida
pir2:T05850	+ 64.50	113.01	36.85	718	! homeobox protein ATML1, L1-spec
pir2:S26317	+ 64.00	125.62	47.73	110	! Ig heavy chain V region - mouse
pir2:C81018	+ 64.00	112.03	42.01	714	! iron-regulated outer membrane p
pir2:A71724	- 63.50	106.26	45.78	1374	! dna-directed RNA polymerase bc
pir2:JE0099	+ 63.00	109.88	54.49	725	! neural cell adhesion molecule 1
pir2:S37216	+ 63.00	108.00	53.54	940	! hexon protein - human adenoviru
pir2:A83652	- 63.00	107.89	53.48	954	! hypothetical protein BH0017 [in
pir1:JUXNL	- 63.00	106.93	53.00	1088	! neural cell adhesion molecule
pir2:PH1561	+ 62.50	120.46	69.24	147	! Ig heavy chain V region (clone
pir2:S69472	- 62.50	120.26	69.11	151	! hypothetical protein YPR053c -
pir2:S29594	+ 62.50	119.07	68.34	178	! Ig gamma chain (WM65) - mouse
pir2:A82822	+ 62.50	108.68	61.98	744	! NADH-ubiquinone oxidoreductase,
pir2:S20582	+ 62.50	107.33	61.21	895	! dystrophin-associated glycoprot
pir2:T21104	- 62.00	106.92	70.15	823	! hypothetical protein F19B2.6 -
pir2:A75455	- 61.50	114.98	87.05	236	! hypothetical protein - Deinococ
pir2:C72321	+ 61.50	109.51	82.69	501	! conserved hypothetical protein
pir2:T51915	+ 61.50	107.10	80.84	698	! hypothetical protein B23111.270
pir2:T46033	- 61.50	104.48	78.86	1002	! receptor protein kinase-like p
pir2:S14201	+ 61.00	102.92	77.72	1242	! probable adenylate/quantilate c
pir2:S26316	+ 61.00	119.64	104.63	108	! Ig heavy chain V region - mouse
pir2:B30515	+ 61.00	118.76	103.76	122	! Ig heavy chain V region (5-IE4)
pir2:T130538	+ 61.00	118.70	103.70	123	! Ig heavy chain V region (174.2E
pir2:H72585	+ 61.00	114.81	99.98	210	! hypothetical protein APEL157 -
pir2:A49505	+ 61.00	112.72	98.09	278	! lectin light chain - Entamoeba
pir2:A48452	+ 61.00	112.52	97.85	288	! galactose-inhibitable lectin 35
pir2:E83488	+ 61.00	111.23	96.67	344	! hypothetical protein PA1255 [in
pir2:T23175	- 61.00	106.60	92.56	650	! hypothetical protein K01C8.3 -
pir2:T00875	+ 61.00	105.48	91.59	759	! hypothetical protein At2g45620

pir2:T06246 - 61.00 104.17 90.47 909 ! aspartate kinase (EC 2.7.2.4
pir2:T06242 - 61.00 104.11 90.42 916 ! aspartate kinase (EC 2.7.2.4
pir1:QBE47 + 61.00 102.60 89.14 1128 ! DNA-binding protein - human
pir2:B8431 - 61.00 102.50 89.06 1143 ! probable receptor protein k
seq_name: pir2:JG0168
seq_documentation_block:
C:Species: Mus musculus (house mouse)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
C:Accession: JG0168
R:Komiyama, T.; Tanigawa, Y.; Hirohashi, S.
Biochem. Biophys. Res. Commun. 255, 347-351, 1999
A:Title: Cloning and identification of the gene gob-5, which is expressed in intestine
A:Reference number: JG0168; MUID:99160866
A:Accession: JG0168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-913 <KOM>
A:Cross-references: DDBJ:AB016592
alignment_scores:
Quality: 319.00 Length: 76
Ratio: 4.493 Gaps: 0
Percent Similarity: 93.421 Percent Identity: 76.316
alignment_block:
US-09-049-696-10 x JG0168
Align seg 1/1 to: JG0168 from: 1 to: 913
1 GGACAGTATCTGGACAGCAGCCGTGGGAAAGACACTTTGTTCTTAT 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
505 GlySerValIleValAspSerValGlyLysAspThrLeuPheIleu 521
51 CACCTGGACAAACGACGCTTGTAGTGGACAAACACCAAAATGGCGCTACCTC 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
521 eThrTrpThrHisProProThrIlePheIleTrpAspProSerGlyV 538
101 AGAAGCAAGTGGCTTGTAGTGGACAAACACCAAAATGGCGCTACCTC 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
538 aIguInasnglyPheIleLeuAspThrThrLysValAlaTyrLeu 554
151 CAAATCCAGGATTCCTAAGGTGGACCTTGGAAATACAGTCTGCAAGC 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
555 GlnValProGlyThrAlaLysValGlyPheTrpLysTyrSerIleGlnAl 571
201 AGCTCACAAACCTTGACCTGACTGTC 228
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
571 aSerSerGlnThrLeuThrLeuThrVal 580
seq_name: pir2:T02205
seq_documentation_block:
Lu-ECAM-1 protein - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 01-Dec-2000
C:Accession: T02205; T02152; T02171
R:Elble, R.C.; Widom, J.; Gruber, A.D.; Abdel-Ghany, M.; Levine, R.; Goodwin, A.; Pau
submitted to the EMBL data library, April 1997
A:Description: Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial
A:Reference number: Z14590
A:Accession: T02205
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-905 <ELB>
A:Cross-references: EMBL:AF001261; NID:g2623762; PIDN:AAB86529.1; PID:g2623763
A:Experimental source: lung
A:Accession: T02152
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA

alignment_block:
US-09-049-696-10/rev x T45935

172 GTTGGCACTTGGAAATAC 189
::: ||::: :::::
126 AlaTrpThrTyrArgHis 131

Align. seq 1/1 to: E64860 from: 1 to: 263

alignment block:

US-09-049-696-10/rev x T45935

F;19-88/Domain: immunoglobulin homology <IMM>

alignment_scores:
Quality: 66.00 Length: 76
Ratio: 1.692 Gaps: 6
Percent Similarity: 51.316 Percent Identity: 35.526

alignment_block:
US-09-049-696-10 x PH1215 ..

Align seg 1/1 to: PH1215 from: 1 to: 227

```
12 CGTGACAGCAGCCGCGGAGAGGACACTTCTTCTTATCATCTGGACAA 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
75 ArgAlaGluTrpGluGlnLysAspGluPheIleCysArgAlaValHisGI 91
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
62 CGCAGCCTCCCA.....TCCAGTGGACAGAGCAAGTGGCTT 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
91 uAlaAspSerProSerGlnThrValGlnArgAlaValSerValAsnProG 108
82 CTCGGGA.....TCCAGTGGACAGAGCAAGTGGCTT 116
:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
108 lYLeuAlaGlyGlySerAlaGlnSerGlnArgAlaProAspArgValLeu 124
|||..|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 TGT...ACTGGACAAACACCAAAATGGCTTACT..... 149
|||..|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 CysHisSerGlyGln.....GlnGlnGlyLeuProArgAlaAlaGlyGI 139
150 .....CCAAATCCAGG...CATTCG 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
139 ySerValProHisProArgCysHisCys 148
```

seq_name: pir4:A47271

seq_documentation_block:
nitrophenyl phosphonate-specific antibody 48G7 heavy chain VDJ - synthetic (fragment)
C:Species: synthetic
A:Note: Mus musculus (house mouse) gene engineered and expressed in Escherichia coli
C:Date: 21-Sep-1993 #sequence_revision 11-Aug-1995 #text_change 11-Aug-1995
C:Accession: A47271
R:Lesley, S.A.; Patten, P.A.; Schultz, P.G.
Proc. Natl. Acad. Sci. U.S.A. 90, 1160-1165, 1993
A:Title: A genetic approach to the generation of antibodies with enhanced catalytic acti
A:Reference number: A47271; MUID:93165660
A:Accession: A47271
A:Molecule type: DNA; protein
A:Residues: 1-114 <LES>
A:Note: sequence extracted from NCBI backbone (NCBIN:124854, NCBI:124855)
A:Note: parts of this sequence were determined by protein sequencing
F;22-96/Disulfide bonds: #status predicted

alignment_scores:
Quality: 65.50 Length: 85
Ratio: 1.560 Gaps: 4
Percent Similarity: 49.412 Percent Identity: 29.412

alignment_block:
US-09-049-696-10 x A47271 ..

Align seg 1/1 to: A47271 from: 1 to: 114

```
31 AAGGACACTTCTTCTTATCATCTGGACAGCAG...CTTCCCAAT 77
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30 LysAspThr.....TyrMethIstPValLysGlnArgProLysGlnGI 44
78 CCTTCTCTGG.....G 88
||| |||
44 yLeuGluTrpIleGlyArgIleAspProAlaAsnValAspThrLysTyra 61
89 ATCCCAAGTGGACAGCAAGTGGCTTGTAGTGCACAAACACCAA 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
61 sPProLysPheGlnAspLysAlaThrIleThrAlaAspThrSerSerLys 77
139 ATGGCTTACCTCCAAATCCAGGATTCCTAAGTTGGCCTTGGAAATA 188
78 ThrThrTyrlLeuGlnLeuSerGlyLeuThrSerGluAspThrAlaValty 94
189 CAGTCTGCAAGCAAGC.....TCAACAACTTGACCTGA 223
94 rTyrcysAlaSerTyrlGlyIleTyrlTrpGlyGlnGlyThrThrLeut 111
224 CTGTC 228
|||||
111 hrVal 112
```

seq_name: pir2:S20901

seq_documentation_block:

titin - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999
C:Accession: S20901; I46520
R:Labeit, S.; Gautel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A:Title: Towards a molecular understanding of titin.
A:Reference number: S20897; MUID:92258380
A:Accession: S20901
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-6805 <LAB>
A:Cross-references: EMBL:X64696
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992
R:Labeit, S.; Barlow, D.P.; Gautel, M.; Gibson, T.; Holt, J.; Hsieh, C.L.; Francke, U
Nature 345, 273-276, 1990
A:Title: A regular pattern of two types of 100-residue motif in the sequence of titin
A:Reference number: I46520; MUID:90238553
A:Accession: I46520
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 4235-5250 <LA2>
A:Cross-references: EMBL:X17329; NID:gl756; PIDN:CAA35207.1; PID:g930251
C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology;
C:Keywords: muscle

alignment_scores:
Quality: 65.50 Length: 81
Ratio: 1.560 Gaps: 4
Percent Similarity: 51.852 Percent Identity: 25.926

alignment_block:
US-09-049-696-10 x S20901 ..

Align seg 1/1 to: S20901 from: 1 to: 6805

```
7 GTGATCTGCGACAGCACCGTGGAAAGGACACTTCTTCTT..... 48
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5957 ValLysLeuGluAsnSerSerGlyThrLysThrAlaPheValThrVally 5973
49 ATCACTTGGACAGCGACGCTCCCAA..... 75
5973 sValLeuAspThrProGlyProGlnAsnLeuAlaValLysGluValL 5990
76 .....ATCCTTCTCTGGATGCC.....ACTGA 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5990 ysLysAspSerAlaValLeuValTrpGluProIleAspGlyGly 6006
100 CAGACCAAGTGGCTTCTACTGGACAAA.....AACACCAAAATGCC 143
6007 AlaLysValArgAsnTrpValIleAspLysArgGluSerThrArgLysAl 6023
144 CTACCTCCAAATCCAGGATTCCTAAGTTGGCCTTGGAAA 186
6023 atyAlaAsnValSerSerLysCysAsnLysThrThrPheLys 6037
```

seq_name: pir2:F64860

```

seq_documentation_block:
  Hypothetical protein bll53 - Escherichia coli (strain K-12)
  C:Species: Escherichia coli
  C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
  C:Accession: F64860
  R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; B
    .; Rose, D.J.; Mau, B.; Shao, Y.
    Science 277, 1453-1462, 1997
  A:title: The complete genome sequence of Escherichia coli K-12.
  A:Reference number: A64720; MUID:97426617
  A:Accession: F64860
  A:Status: preliminary; nucleic acid sequence not shown; transla
  A:Molecule type: DNA
  A:Residues: 1-194 <BLAT>
  A:Cross-references: GB:AE000214; GB:U000096; NID:gl787382; PIDN:
  A:Experimental source: strain K-12, substrain MG1655

```

```

alignment_scores:
  Quality: 65.00      Length: 67
  Ratio: 2.241       Gaps: 3
  Percent Similarity: 43.284  Percent Identity: 31.343

alignment_block:
  US-09-049-696-10/rev x F64860 ..
  Align seg 1/1 to: F64860 from: 1 to: 194

```

Align seg 1/1 to: F64860 from: 1 to: 194

209 TGTGAGCTTGCTTGC.....AGACTGTA 187
 |||:::|||||
 126 CysSerSerAlaCysThrAspAlaValasAlaProGluTrpArgTyr 142
 |||:::|||||
 186 TTTCCAAGTGCACAACCTTAGCAATGCCTGGGATTGGAGCTAGGCCATT 137
 |||:::|||||
 142 rlrpGlnVal.....AsnMetProAlaThr..... 150
 136 TGGTGTTTTTGTCCACTACAAGCCACCTTGCTTCTGTCCACTGGGATCC 87
 :::::|||||
 151ThrAsnSerThrTrpMetThrCysGlyAspProCysAspSer 164
 86 CAGAGAAGGATTGGGAGGCTGCTGTCTCCAGGTGATAAGAACAAAGT 37
 |||:::|||||
 165 AlaLeuArgIleTrpGlyAspThrValValGluCysValLeuAsnLysLe 181
 |||:::|||||
 36 G 36
 181 : 181

seq_name: pir2:JE0100

seq_documentation_block:
neural cell adhesion molecule 2 - African clawed frog
N;Alternate names: N-CAM 2
C;Species: Xenopus laevis (African clawed frog)
C;Date: 19-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C;Accession: JE0100
R;Kudo, M.; Takayama, E.; Tadakuma, T.; Shiokawa, K.
Biochem. Biophys. Res. Commun. 245, 127-132, 1998
A;Title: Molecular cloning of ssd-form neural cell adhesion molecules (N-CAMs) as the ma
A;Reference number: JE0099; MUID:98204770
A;Accession: JE0100
A;Molecule type: mRNA
A;Residues: 1-725 <KUD>
A;Cross-references: DDBJ:AB008163; NID:g3116228; PIDN:BAA25932.1; PID:g3116229
A;Experimental source: heart
C;Comment: This protein mediates and regulates various cell-cell interactions through b
C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
F;413-475/Domain: immunoglobulin homology <IMM>
F;512-589/Domain: fibronectin type III repeat homology <3FR>

```

alignment_scores:
    Quality: 65.00      Length: 61
    Ratio:   1.806     Gaps: 2
    Percent Similarity: 59.016      Percent Identity: 32.787

alignment_block:
    US-09-049-696--10 x JE0100 ..

Align seg 1/1 to: JE0100 from: 1 to: 725

22 ACCGTGGGAAGCAGCACTTGTGTTCTWTATCACCTGGACAACGCACCCGCC 71
||| |||:::||:::||:::||:::||:::||:::||:::||:::||:::||:
127 ThrGluglycIuAspAlaValIllelEcysAspValSerSeriLepr 143

72 CCAAATCCCTTCCTCGGATCCCAGTGGA.....CAGAAGC 106
|||||::|::| |::|::|
143 OsErllellieThrTrpArgHisLysGLyLysAspValllePheLysLysA 160

107 AAGGTGGCTTTTGATGGGACAAATAACCACAAAATGGCCTACTCCAATC 156
:: |||||::|::|::|::|::|::|::|::|::|
160 spValArgPheValAlValAlaAsnAsn.....TyrlEuclnile 173

157 CCAGGCATTGCTTAAGGTGGCCACTTGGAAATAC 189
||||| |||::| ::|::|
174 ArgGlyilEllysLysThrsAspGluGlyAsnTyr 184

seq_name: pirl.JN0635
```

```

seq_documentation_block:
  neural cell adhesion molecule 2 precursor - African clawed frog
  C:Species: Xenopus laevis (African clawed frog)
  C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change
  C:Accession: JN0635
  R:Tonissen, K.F.; Krieg, P.A.
  Gene 127, 243-247, 1993
  A:Title: Two neural-cell adhesion molecule(NCAM)-encoding genes
  A:Reference number: JN0635; MUID:93273239
  A:Accession: JN0635
  A:Molecule type: mRNA
  A:Residues: 1-1092 <TON>
  A:Cross-references: GB:W76710; NID:Q214611; PIDN:AAA49910.1; PIR
  C:Comment: NCAM mediates cell-cell adhesion via homophilic bind
  C:Genetics:

```

A: Gene: NCAM2
C: Superfamily: neural cell adhesion molecule; fibronectin type
C: Keywords: alternative splicing; cell adhesion; duplication; h
F: 1-109/Domain: signal sequence #status predicted <SIG>
F: 20-192/product: neural cell adhesion molecule 2 #status pred
F: 20-705/Domain: extracellular #status predicted <EXT>
F: 34-95/Domain: immunoglobulin homology <IMM1>
F: 129-188/Domain: immunoglobulin homology <IMM2>
F: 149-153/Region: heparin binding #status predicted
F: 158-162/Region: heparin binding #status predicted
F: 317-381/Domain: immunoglobulin homology <IMM3>
F: 413-475/Domain: immunoglobulin homology <IMM4>
F: 512-589/Domain: immunoglobulin type III repeat homology <FN3A>
F: 619-680/Domain: fibronectin type III repeat homology <FN3B>
F: 706-723/Domain: transmembrane #status predicted <TM>
F: 724-1092/Domain: intracellular #status predicted <INT>
F: 419-136-186, 232-282, 323-379, 420-473/Duplicate bonds: #status
F: 219, 310, 341, 417, 443, 472/Binding site: carbohydrate (Asn) (cov

```

alignment_scores:      Quality: 65.00      Length: 61
                       Ratio: 1.806      Gaps: 2
                       Percent Similarity: 59.016      Percent Identity: 32.787

alignment_block:
US-09-049-696-10 x JN0635      ..

Align seg 1/1 to: JN0635 from: 1 to: 1092

```

[illegible]

A:Reference number: Z15454
A:Accession: T05850
A:Molecule type: DNA
A:Residues: 1-718 <BEV>
A:Cross-references: EMBL:AL035527; GSPDB:GN00062; ATSP:F17L22.210
R:Experimental source: cultivar Columbia; BAC clone F17L22
R:Nadeau, J.A.; Lee, E.K.; O'Neill, S.D.
submitted to the EMBL Data Library, October 1995
A:Description: The Arabidopsis homeobox gene A20 is expressed in ovules.
A:Reference number: S71211
A:Accession: S71211
A:Molecule type: mRNA
A:Residues: 1-317, R, 319-483, D', 485-718 <NAD>
A:Cross-references: EMBL:U37589; NID:g1173829; PID:g1881536
C:Genetics:
A:Gene: ATSP:F17L22.210; ATML1
A:Map position: 4
A:Introns: 35/2; 74/3; 293/3; 327/3; 394/3; 453/1; 543/3; 575/3
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:19-75/Domain: homeobox homology <Hox>

alignment_scores:
Quality: 64.50 Length: 76
Ratio: 1.697 Gaps: 3
Percent Similarity: 50.000 Percent Identity: 28.947

alignment_block:
US-09-049-696-10/rev x T05850 ..
Align seg 1/1 to: T05850 from: 1 to: 718

185 TTCCAAGTCCCAACCTTAGCAATGCT..... 159
|||||.....
332 PheGlnValProSerProLeuValProThrArgGluAsnTyrPheValAr 348
158GGGATTTCGAGGTAGCGCAATTTTCGTGT 131
|||||.....
348 gTyrCysLysGlnHisSerAspGlyIleTrpAla...ValValAspValS 364
130 TTTTGTCCACTACAAAGCACCTTGCTCTGTCACCTGGATCCACAGAGA 81
|||||.....
364 erLeuAspSerLeuArgPro.....SerProIleThrArgSerArg 377
80 AGGATTGGGGAGCGTCGTTGCCAGGTGATAAGAAACAAAGTGCCTT 31
|||
378 ArgArgProSerGlyCysLeuIleGlnGluLeuGlnAsnGlyTyrSerLy 394
30 TCCCACGGTGTCTGCCAGTCACTGTG 3
|||
394 sValThrTrpValGluHisIleGluVal 403

seq_name: pir2:S26317

seq_documentation_block:
Ig heavy chain V region - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1998 #sequence_revision 19-Mar-1998 #text_change 21-Jan-2000
C:Accession: S26317
R:Stark, S.E.; Caton, A.J.
J. Exp. Med. 174, 613-624, 1991
A:Title: Antibodies that are specific for a single amino acid interchange in a protein
A:Reference number: S26309; MUID:91341421
A:Accession: S26317
A:Molecule type: mRNA
A:Residues: 1-110 <STA>
A:Cross-references: EMBL:X59186
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:4-87/Domain: immunoglobulin homology <IMM>

alignment_scores:
Quality: 64.00 Length: 90
Ratio: 1.333 Gaps: 4
Percent Similarity: 53.333 Percent Identity: 24.444

alignment_block:
US-09-049-696-10 x S26317 ..
Align seg 1/1 to: S26317 from: 1 to: 110

37 ACTTTGTTTCTTATCACCTGGACAACGCAG...CCTCCCAAAATCCTTCT 83
||| ...
19 ThrSerTyrTrpMetAsnTrpValLysGlnArgProGlyGlnGlyLeuG1 35
84 CTGG.....GATCCCACTGGACAGAAAG..... 105
|||
35 uTrpIleGlyMetIleAspProSerAspSerGluThrHisTyrAsnGlnM 52
106CAAGTGGCTTTGTAGTGACAAAAACACCAAAATGGCC 144
.....
52 etPheArgAspLysAlaThrLeuThrValAspLysSerSerAsnThrAla 68
145 TACCTCCAAATCCAGGCATT..... 165
|||
69 TyrMetGlnLeuSerSerLeuThrSerGluAspSerAlaValTyrTyrCy 85
166GCTAAGTTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCAC 208
.....
85 sAlaArgGlyAsnTyrGlyAsnTyrAlaTyrAlaLeuAspTyrTrpGlyG 102
209 AAACCTTGACCCCTGACTGTC 228
||
102 InGlyThrSerValThrVal 108

This Page Blank (uspto)


```
42 GTTCTTATCATCCCTGGACAGCAG...CCTCCCAAAATCCTTCTCTGG. 87
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
32 rTyrTrpMetAsnTrpValLysGlnArgProGlyGlnGlyLeuGluTrpI 49
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
88 .....GATCCAGTGGACAGAG..... 105
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
49 LeGlyMetIleAspProSerAspSerGluThrGlnTyrAsnGlnMetPhe 65
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
106 .....CAAGTGGCTTGTAGTGGACAAACACCAAAATGGCCTACCT 149
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
66 LysAspLysAlaLeuThrValAspLysSerAsnThrAlaTyrMe 82
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
150 CCAAAATCCAGCGCATGCTAAG..... 171
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
82 tGlnLeuSerSerLeuThrSerGluAspSerAlaValTyrTyrCysAlaL 99
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
172 .....GTTGGCACTTGGAAATACAGTCTGCAGCAGCAAGCTCACA 210
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
99 ysGlyGlyAlaSerGlyAspTrpTyrPheAspValTrpGlyGlnGlyThr 115
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
211 ACCTTGACCTGACT 225
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
116 ThrValThrValSer 120
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

seq_name: /cgn2_6/ptodata/2/iaa/5a_COMB.pep:us-08-465-473B-11

seq_documentation_block:
; Sequence 11, Application US/08465473B
; Patent No. 5939531
; GENERAL INFORMATION:
; APPLICANT: Wells, Winfried S.
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5939531man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; TITLE OF INVENTION: Growth Factor Receptor
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NOVARTIS Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901-6940
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,473B
; FILING DATE: 5 June 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pfeiffer, Hessa J.
; REGISTRATION NUMBER: 22,640
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)522 6940
; TELEFAX: (908)522 6955
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 241 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

```
MOLECULE TYPE: protein
US-08-465-473B-11
Length: 105
Gaps: 5
Ratio: 1.196
Percent Similarity: 53.333
Percent Identity: 22.857

alignment_scores:
Quality: 67.00
Ratio: 1.196
Percent Similarity: 53.333
Percent Identity: 22.857

alignment_block:
US-09-049-696-10 x US-08-465-473B-11
Align seg 1/1 to: US-08-465-473B-11 from: 1 to: 241
1 GGCACAGTGTATCGTGGACAGCACCGTGGGAAGGAC.....ACTTT 41
||||| : : : : : : : : : : : : : : : : : : : : : : : :
16 GlyThrSerValLysLeuSerCysLysAlaSerAspTyrThrPheThrse 32
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
42 GTTCTTATCATCCCTGGACAGCAG...CCTCCCAAAATCCTTCTCTGG. 87
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
32 rTyrTrpMetAsnTrpValLysGlnArgProGlyGlnGlyLeuGluTrpI 49
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
88 .....GATCCAGTGGACAGAG..... 105
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
49 LeGlyMetIleAspProSerAspSerGluThrGlnTyrAsnGlnMetPhe 65
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
106 .....CAAGTGGCTTGTAGTGGACAAACACCAAAATGGCCTACCT 149
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
66 LysAspLysAlaLeuThrValAspLysSerAsnThrAlaTyrMe 82
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
150 CCAAAATCCAGCGCATGCTAAG..... 171
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
82 tGlnLeuSerSerLeuThrSerGluAspSerAlaValTyrTyrCysAlaL 99
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
172 .....GTTGGCACTTGGAAATACAGTCTGCAGCAGCAAGCTCACA 210
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
99 ysGlyGlyAlaSerGlyAspTrpTyrPheAspValTrpGlyGlnGlyThr 115
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
211 ACCTTGACCTGACT 225
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
116 ThrValThrValSer 120
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

seq_name: /cgn2_6/ptodata/2/iaa/5a_COMB.pep:us-08-235-838-16

seq_documentation_block:
; Sequence 16, Application US/08235838
; Patent No. 5571894
; GENERAL INFORMATION:
; APPLICANT: Wells, Winfried S.
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5571894man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; TITLE OF INVENTION: Growth Factor Receptor
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235,838
; FILING DATE: TBA
; CLASSIFICATION: 435
```

```

; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-235-838-16

alignment_scores:
    Quality: 67.00      Length: 105
    Ratio: 1.196       Gaps: 5
    Percent Similarity: 53.333      Percent Identity: 22.857

alignment_block:
US-09-049-696-10 x US-08-235-838-16
..
Align seg 1/1 to: US-08-235-838-16 from: 1 to: 637

1  GCACAGTATCATCTGGACAGCACCGTGGAAAGGAC.....ACTTT 41
   ||||| : : : ||| : : : |||
47  GlyThrSerValLysLeuSerCysLysAlaSerAspTyrThrPheThrSe 63
42  GTTCTTATCATCTGGACAGCAG...CCTCCCAAAATCCTTCTCTGG.. 87
   : : : : : ||||| : : ||| ||| ||| |||
63  rTyrTrpMetAsnTrpValLysGlnArgProGlyGlnGlyLeuGluTrpI 80
88  .....GATCCACGTGGACAGAAG..... 105
80  leGlyMetIleAspProSerAspSerGluThrGlnTyrAsnGlnMetPhe 96
106 .....CAAGTGGCTTTGTAGTGGCAAAACACCAAAATGCCTACCT 149
   : : : : : ||||| : : ||| ||| ||| |||
97  LysAspLysAlaAlaLeuThrValAspLysSerAsnThrAlaTyrMe 113
150  CCAAAATCCAGGCAATGCTAAG..... 171
113  tGlnLeuSerSerLeuThrSerGluAspSerAlaValTyrTyrCysAlaL 130
172 .....GTTGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAA 210
   ||| ||| : : : : :
130  ysglyGlyAlaSerGlyAspTrpTyrPheAspValTrpGlyGlnGlyThr 146
211  ACCTTACCTCACT 225
   ||| : : : : :
147  ThrValThrValSer 151

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-465-473B-16

seq_documentation_block:
; Sequence 16, Application US/08465473B
; Patent No. 5939531
; GENERAL INFORMATION:
; APPLICANT: Wels, Winfried S.
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5939531man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a

```

```

; TITLE OF INVENTION: Growth Factor Receptor
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NOVARTIS Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901-6940
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,473B
; FILING DATE: 5 June 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pfeiffer, Hesna J.
; REGISTRATION NUMBER: 22,640
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)522 6940
; TELEFAX: (908)522 6955
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-465-473B-16

alignment_scores:
    Quality: 67.00      Length: 105
    Ratio: 1.196       Gaps: 5
    Percent Similarity: 53.333      Percent Identity: 22.857

alignment_block:
US-09-049-696-10 x US-08-465-473B-16
..
Align seg 1/1 to: US-08-465-473B-16 from: 1 to: 637

1  GCACAGTATCATCTGGACAGCACCGTGGAAAGGAC.....ACTTT 41
   ||||| : : : ||| : : : |||
47  GlyThrSerValLysLeuSerCysLysAlaSerAspTyrThrPheThrSe 63
42  GTTCTTATCATCTGGACAGCAG...CCTCCCAAAATCCTTCTCTGG.. 87
   : : : : : ||||| : : ||| ||| ||| |||
63  rTyrTrpMetAsnTrpValLysGlnArgProGlyGlnGlyLeuGluTrpI 80
88  .....GATCCACGTGGACAGAAG..... 105
80  leGlyMetIleAspProSerAspSerGluThrGlnTyrAsnGlnMetPhe 96
106 .....CAAGTGGCTTTGTAGTGGCAAAACACCAAAATGCCTACCT 149
   : : : : : ||||| : : ||| ||| ||| |||
97  LysAspLysAlaAlaLeuThrValAspLysSerAsnThrAlaTyrMe 113
150  CCAAAATCCAGGCAATGCTAAG..... 171
   : : : : :
113  tGlnLeuSerSerLeuThrSerGluAspSerAlaValTyrTyrCysAlaL 130
172 .....GTTGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAA 210
   ||| ||| : : : : :
130  ysglyGlyAlaSerGlyAspTrpTyrPheAspValTrpGlyGlnGlyThr 146

```

211 ACCTTGACCTGACT 225
||||:||||:||||:
147 ThrValThrValSer 151

seq_name: /cgn2_6/ptodata/2/iaa/5b_comb.pep:us-08-800-198-2

seq_documentation_block:
; Sequence 2, Application US/08800198
; Patent No. 5942602
; GENERAL INFORMATION:
; APPLICANT: WELLS, WINFRIED S.
; APPLICANT: SCHMIDT, MATHIAS
; APPLICANT: VAKALOPOULOU, EVANGELIA
; APPLICANT: SCHNEIDER, DOUGLAS
; TITLE OF INVENTION: GROWTH FACTOR RECEPTOR ANTIBODIES
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MILLEN, WHITE, ZELANO & BRANIGAN, P.C.
; STREET: 2200 CLARENDON BLVD. SUITE 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: US
; ZIP: 22201

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,198
; FILING DATE: 13-FEB-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: HAMLET-KING, DIANA
; REGISTRATION NUMBER: 33,302
; REFERENCE/DOCKET NUMBER: SCH 1576
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-243-6333
; TELEFAX: 703-243-6410
; INFORMATION FOR SEQ ID NO: 2:
; LENGTH: 119 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; US-08-800-198-2

alignment_scores:
Quality: 65.00 Length: 88
Ratio: 1.354 Gaps: 5
Percent Similarity: 54.545 Percent Identity: 26.136

alignment_block:

US-09-049-696-10 x US-08-800-198-2 ..
Align seg 1/1 to: US-08-800-198-2 from: 1 to: 119

37 ACTTTGTTTCTTATCACCTGGACACGCGAG...CCTCCCCAAATCCTTCT 83
||| :||:||||| ||| |||:||||: |||
30 ThrAsnTyrTrpMetHisTrpValThrGlnArgProGlyGlnValLeuVa 46
84 CTGG.....GATCCAGT.....GGACAGA 103
||| :||:||||| ||| |||:||||: |||
46 lTrpIleGlyThrAsnProAsnThrGlyTyrThrAspPheAsnGlnL 63
104 AG.....CAAGGTGGCTTTGTAGTGGACAAACACCAAAATGGCC 144
|| :||:||||| ||| |||:||||: |||
113 hrThrValThrVal 117

63 ysPheLysAspLysAlaThrLeuThrAlaAspLysSerSerSerThrAla 79
145 TACCTCCAAATCCAGGCATT..... 165
|||:||||: |||:
80 TyrMetGlnLeuSerGlyLeuThrSerGluAspSerAlaValTyrTyrCy 96
166 GCTAAGGTGGCACTTGGAAATACAGTCTGCAAGCAGCTCACAAACCT 214
|||: |||: |||: |||: |||: |||: |||:
96 sAlaArgGlyAspTyrTyrGlyTyrAspPheAlaTyrTrpGlyGlnGlyT 113
215 TGACCTGACTGTC 228
|||:||||| |||:
113 hrThrValThrVal 117

seq_name: /cgn2_6/ptodata/2/iaa/6A_comb.pep:us-09-296-595-2

seq_documentation_block:
; Sequence 2, Application US/09296595A
; Patent No. 6129915
; GENERAL INFORMATION:
; APPLICANT: WELLS, WINFRIED S.
; APPLICANT: SCHMIDT, MATHIAS
; APPLICANT: VAKALOPOULOU, EVANGELIA
; APPLICANT: SCHNEIDER, DOUGLAS
; TITLE OF INVENTION: GROWTH FACTOR RECEPTOR ANTIBODIES
; FILE REFERENCE: SCH-1576 D1
; CURRENT APPLICATION NUMBER: US/09/296,595A
; EARLIER FILING DATE: 1999-04-23
; EARLIER APPLICATION NUMBER: 08/800,198
; EARLIER FILING DATE: 1997-02-13
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Murine sp.
; US-09-296-595-2

alignment_scores:
Quality: 65.00 Length: 88
Ratio: 1.354 Gaps: 5
Percent Similarity: 54.545 Percent Identity: 26.136

alignment_block:

US-09-049-696-10 x US-09-296-595-2 ..
Align seg 1/1 to: US-09-296-595-2 from: 1 to: 119

37 ACTTTGTTTCTTATCACCTGGACACGCGAG...CCTCCCCAAATCCTTCT 83
||| :||:||||| ||| |||:||||: |||
30 ThrAsnTyrTrpMetHisTrpValThrGlnArgProGlyGlnValLeuVa 46
84 CTGG.....GATCCAGT.....GGACAGA 103
||| :||:||||| ||| |||:||||: |||
46 lTrpIleGlyThrAsnProAsnThrGlyTyrThrAspPheAsnGlnL 63
104 AG.....CAAGGTGGCTTTGTAGTGGACAAACACCAAAATGGCC 144
|| :||:||||| ||| |||:||||: |||
63 ysPheLysAspLysAlaThrLeuThrAlaAspLysSerSerSerThrAla 79
145 TACCTCCAAATCCAGGCATT..... 165
|||:||||: |||:
80 TyrMetGlnLeuSerGlyLeuThrSerGluAspSerAlaValTyrTyrCy 96
166 GCTAAGGTGGCACTTGGAAATACAGTCTGCAAGCAGCTCACAAACCT 214
|||: |||: |||: |||: |||: |||: |||:
96 sAlaArgGlyAspTyrTyrGlyTyrAspPheAlaTyrTrpGlyGlnGlyT 113
215 TGACCTGACTGTC 228
|||:||||| |||:
113 hrThrValThrVal 117

APPLICANT: SCHLEGEL, C. Richard
APPLICANT: JENSON, A. Bennett
APPLICANT: GHIM, Shin-Je
TITLE OF INVENTION: PAPILLOMAVIRUS VACCINES
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,281
FILING DATE: 01-OCT-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/004,691
FILING DATE: 02-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 010091-015
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-724-281-2

alignment_scores:
Quality: 65.00 Length: 26
Ratio: 3.421 Gaps: 2
Percent Similarity: 73.077 Percent Identity: 50.000

alignment_block:
US-09-049-696-10 x US-08-724-281-2 ..
Align seg 1/1 to: US-08-724-281-2 from: 1 to: 503

78 CCTTCTCTGGATCCCTGGACAGACAGCAAGTGGCTTTGTAGTGGACA 127
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
166 ProAlaLeuGluHisTrpSerLeuThrArgTrp...CysThrGlyG1 181
128 AAACACCAAAATGCG...CTACCTCCA 152
181 nValHisThrAlaGlyGlnCysProPro 190

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-560-398-12
seq_documentation_block:
Sequence 12, Application US/08560398
Patent No. 5907082
GENERAL INFORMATION:
APPLICANT: O'Neill, Sharman
APPLICANT: Nadeau, Jeanette
TITLE OF INVENTION: Ovule-Specific Gene Expression
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco

STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/560,398
FILING DATE: 17-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-063300US
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-560-398-12

alignment_scores:
Quality: 64.50 Length: 76
Ratio: 1.697 Gaps: 3
Percent Similarity: 50.000 Percent Identity: 28.947

alignment_block:
US-09-049-696-10/rev x US-08-560-398-12 ..
Align seg 1/1 to: US-08-560-398-12 from: 1 to: 718

185 TTCCAAGTGCCACCTTAGCAATGCCT..... 159
|||||:|||||:|||||:|||||:|||||:|||||:|||||
332 PheGlnValProSerProLeuValProThrArgGluAsnTyrPheValar 348
158GGATTGGAGGTAGGCCATTTGGTGT 131
348 gTyrCysLysGlnHisSerAspGlyIleTrpAla...ValValAspValS 364
130 TTTTGTCCACTACAAAGCCACTTGTCTGTCCACTGGGATCCAGACA 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||
364 erLeuAspSerLeuArgPro.....SerProIleThrArgSerArg 377
80 AGGATTTGGGAGGCTGGTGTCTCCAGGTGATAGAAACAAGTCTCTT 31
|||||:|||||:|||||:|||||:|||||:|||||:|||||
378 ArgArgProSerGlyCysLeuIleGlnGluLeuGlnAsnGlyTyrSerLy 394
30 TCCACAGGTGCTGTCCAGCATCACTGTG 3
394 sValThrTrpValGluHisIleGluVal 403

seq_name: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep:PCR-US94-03744-4
seq_documentation_block:
Sequence 4, Application PC/TUS9403744
GENERAL INFORMATION:
APPLICANT: PETRI, WILLIAM A.
APPLICANT: MCCOY, JAMES J.
TITLE OF INVENTION: 35/31 KDA SUBUNIT OF THE
TITLE OF INVENTION: ENTAMOREA
TITLE OF INVENTION: HISTOLYTICA ADHERENCE LECTIN
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 2000 Pennsylvania Avenue, Suite 5500

```
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/03744
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/045,679
; FILING DATE: 09-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 29148-20005.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 288 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-03744-4

alignment_scores:
  Quality: 63.00      Length: 81
  Ratio: 1.969       Gaps: 3
Percent Similarity: 39.506 Percent Identity: 29.630

alignment_block:
US-09-049-696-10 x PCT-US94-03744-4
..
Align seg 1/1 to: PCT-US94-03744-4 from: 1 to: 288

46 CTTATCCTGGACACGAGCTCCCAATCCTTCTGGGATCCAG 95
||||| |||: |||: |||: |||: |||: |||: |||
170 LeullelrlrpsrlysnProPro.....LeuSerAspAlaPr 183
96 TGGACAGACGAGTGGCTTT..... 117
||| |||: |||
183 OGlyThrLysAsnGlnArgPheValtyrValHisProTyrSerThrAspT 200
118 .....GTA 120
200 rPtyrProGluTyrHisSerLysValLysTyrAsnAsnLysGlyArgVal 216
121 GTGCACAAACACCAAAATGGCTACCTCCAAATCCAGGCATTCTAA 170
|||: |||: |||: |||: |||: |||: |||: |||
217 ValGluLysThrLeuGluTrpProThrTyrLysArgHisPheTyrLeup 233
171 GGTGGCATTGGAAATACAGTCTGCAAGCAAGCTCACAAA 211
||| |||: |||: |||: |||: |||: |||
233 roTyrArgLeuAspValAspLeuCysTyrGlnAlaArgLys 246

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-461-697-441

seq_documentation_block:
; Sequence 441, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth

; CITY: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461.697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 441
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-461-697-441

alignment_scores:
  Quality: 62.00      Length: 72
  Ratio: 1.676       Gaps: 3
Percent Similarity: 51.389 Percent Identity: 30.556

alignment_block:
US-09-049-696-10/rev x US-09-461-697-441
..
Align seg 1/1 to: US-09-461-697-441 from: 1 to: 102

195 CACAGTGTATTTCCAAAGTCCCAACCTTAGCAATGCTGGGATTTGGAGGT 146
||| ||| :|||: |||: |||: |||: |||: |||
7 GlnArgValLysAlaGlnThrArgLeuGlyThrArgLeuAspLeuAspTh 23
145 AGCCCATTTTGGTGTGTTTGTCCACTACAAAGCCACCTTG..... 106
|||: |||: |||: |||: |||: |||: |||
23 rAlaHis.....IlePheCysGlnTrpGlnSerCysLeuGlnMetGlyM 38
105 .....CTTCTGTCCACTGGGATCCACAGAGAGGATTTGGGG 70
||||| |||: |||: |||: |||: |||: |||: |||
38 etTyrLeuAsnGlnLeuLeuSerThrProLeuProGluProAspLeuThr 54
63 AGGCTCGTGTTCAGGTGATAAGAAACAAAGTGCTCTTCCACCGCTGC 20
||||| |||: |||: |||: |||: |||: |||
55 ArgLeu.....TyrSerGlySerLe 61
19 TGTCCACGATCACTGT 4
||||| |||
61 uValHisGlyLeuCys 66

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-290-592E-18

seq_documentation_block:
; Sequence 18, Application US/08290592E
; Patent No. 5824307
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, L.
; TITLE OF INVENTION: Human Murine Chimeric Antibodies Against
; TITLE OF INVENTION: Respiratory Syncytial Virus
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELIA, BYRNE, BAIN, GILFILLAN, CECCHI, STEWART &
; ADDRESSEE: OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,592E
```

```

; FILING DATE: August 15, 1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/813,372
; FILING DATE: December 23, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Olstein, Elliot M.
; REGISTRATION NUMBER: 24,025
; REFERENCE/DOCKET NUMBER: 69201-257
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 AMINO ACIDS
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-290-592E-18

alignment_scores:
    Quality: 62.00      Length: 85
    Ratio: 1.409        Gaps: 3
    Percent Similarity: 51.765    Percent Identity: 24.706

alignment_block:
US-09-049-696-10 x US-08-290-592E-18 ..
Align seg 1/1 to: US-08-290-592E-18 from: 1 to: 117

43 TTTCTTATCATCCGCGCAACGCAG...CCTCCCCAAATCCTTCTCTGG... 87
   ::::: ||| |||::: ||| ||| ||| ||| ||| |||
32 TyrTyrlleTyrrpVallysGlnArgProGluGlnGlyLeuGluTrpIli 48
88 .....GATCCCAAGTGAGC 100
48 eglyTrpIleAspProGluAsnGlyAsnThrValPheAspProIysPheG 65
101 AGAAGCAAGGTGGCTTGTAGTGGACAAAACACCAAAATGGCGCTACCTC 150
   || ::::: ||| ::::: ||| |||::: ||| ||| |||
65 lngLylysAlaSerIleThrSerAspThrSerSerAsnThrAlaTyrlleu 81
151 CAATCCCGAGCGAATT.....GCTAA 170
   ||::: :::::
82 GlnLeuSerSerLeuThrSerGluAspThrAlaValTyrrTyrrCysAlaty 98
171 GGTGGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAAACCTTGACCC 220
98 rTyrglyThrSerSerPheAspPheTrpGlyGlnGlyThrThrLeuThrV 115
221 TCAGT 225
   :::::
115 alser 116

seq_name: /cgn2_6/ptodata/2/laa/PCTUS_COMB.pep:PCT-US95-10053-15
seq_documentation_block:
; Sequence 15, Application PC/TUS9510053
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, L.
; TITLE OF INVENTION: Human Murine Chimeric Antibodies Again
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE

```



```

alignment_scores:
    Quality: 62.00      Length: 85
    Ratio: 1.409       Gaps: 3
    Percent Similarity: 51.765   Percent Identity: 24.706

alignment_block:
US-09-049-696-10 x PCT-US96-09448-18 ..

Align seg 1/1 to: PCT-US96-09448-18 from: 1 to: 117

43 TTTCTTTATCACCTGGACAAAGCCAG...CTCCCCCAAAATCCTTCTCTGG... 87
   ::  ||  |||||::  |||  |||  |||  |||  |||  |||  |||  |||
32 TyrTyrlleTyrrpValLysGlnAArgProGluGlnGlyLeuGluTrpI 48

88 .....GATCCACGTGGAC 100
   |||||::  |
48 eGlyTrpIleAspProGluAsnGlyAsnThrValPheAspProLysPheG 1
   |||||::  |
101 AGAAGCAAGGTGGCTTTGTAGTGGCAAAAAACCAACCAATGGCTACCTC 150
   ||  |||||::  |||  :::::  |||||::  |||||::  |||||::
65 lnGlyLysAlaSerIleThrSerAspThrSerSerAsnThrAlaTyrl 81

151 CAATCCCGGCATT.....GCTAA 170
   ||||::  :::::  ||||
82 GlnLeuSerSerLeuthrSerGluAspThrAlaValTyrrCysAlaTy 98
   |||||  :::::  :::::  :::::  :::::  |||||::  |||||::
171 GGTTCGCACTTGGAAATACAGTCTGCAAGCAAGCTCACAACCTTCACCC 220
   |||||  :::::  :::::  :::::  :::::  |||||::  |||||::
98 rTyGlyThrSerSerPheaspPheTrpGlyGlnGlyThrThrLeuThrV 115

221 TGACT 225
   :::::
115 alser 116

```

This Page Blank (uspto)

DR N-PSDB; AAF81927.
XX New nucleic acid encoding calcium activated chloride channel, used to
PT identify, e.g. specific modulators for treating atopic allergy
XX
XX
PS Claim 11; Fig 4B; 75pp; English.
XX
CC The present sequence represents the human Interleukin 9 (IL-9) induced
CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
CC have anti-allergic, anti-asthmatic, anti-inflammatory and
CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
CC used to alleviate asthma (or more generally atopic allergy), while those
CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
CC (increased levels) or IBD (reduced levels), also for monitoring
CC treatment of these conditions. The ICACC proteins can be used:
CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
CC reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
CC identify modulators and binding partners. ICACC polynucleotides can be
CC used to generate transgenic animals or recombinant cells, used to screen
CC for antagonists, also as a source of therapeutic antisense agents or
CC diagnostic probes (for quantifying mRNA expression, e.g. for
CC identification of modulators).
XX
SQ Sequence 914 AA;

alignment_scores:
Quality: 396.00 Length: 76
Ratio: 5.211 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-10 x AAB74824

Align seg 1/1 to: AAB74824 from: 1 to: 914

1 GGACAGTGTATCGTGGACACACCGTGGGAAGGACACTTTGTTCTTAT 50
|||||
504 GlyThrValIleValAspSerThrValGlyLysAspThrLeuPheLeuII 520
51 CACCTGGACAAACGACCTCCCAATCTCTCTGGGATCCAGTGGAC 100
|||||
520 eThrTrpThrThrGlnProGlnIleLeuLeuTrpAspProSerGlyG 537
101 AGAAGCAAGCTGGCTTGTAGTGGACAAACACCAAAATGGCCTACCTC 150
|||||
537 InLysGlnGlyGlyPheValValAspLysAsnThrLysMetAlaTyrLeu 553
151 CAAATCCCAGGCATTCCTAAGGTGGCATTGGCAATATACAGTCTGCAAGC 200
|||||
554 GlnIleProGlyIleAlaLysValGlyThrTrpLysTyrSerLeuGlnAl 570
201 AGCTCACAACCTTGACCTGACTGTC 228
|||||
570 aSerSerGlnThrLeuThrLeuThrVal 579

seq_name: /SID2/gcgdata/geneseq/geneseq/AA2001.DAT:AA24514

seq_documentation_block:

ID AAM24514 standard; Protein; 914 AA.
XX
AC AAM24514;
XX
DT 12-OCT-2001 (first entry)
XX
XX C902P predicted amino acid sequence.
DE Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW immunogenic; gene therapy; vaccine; colon cancer; colon tumour;
XX

OS Homo sapiens.
XX
PN WO200149716-A2.
XX
PD 12-JUL-2001.
XX
XX 29-DEC-2000; 2000WO-US35596.
XX
PR 30-DEC-1999; 99US-0476296.
PR 10-JAN-2000; 2000US-0480321.
PR 15-FEB-2000; 2000US-0504629.
PR 06-MAR-2000; 2000US-0519444.
PR 19-MAY-2000; 2000US-0575251.
PR 29-JUN-2000; 2000US-0609448.
PR 28-AUG-2000; 2000US-0649811.
XX
PA (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
DR WPI; 2001-441847/47.
XX
XX Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer -
XX
PS Claim 2; Page 440-443; 472pp; English.
XX
XX The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be
CC used to treat disorders associated with decreased expression by
CC rectifying mutations or deletions in a patient's genome that affect the
CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host
CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.
XX
SQ Sequence 914 AA;

alignment_scores:
Quality: 396.00 Length: 76
Ratio: 5.211 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-10 x AAM24514

Align seg 1/1 to: AAM24514 from: 1 to: 914

1 GGACAGTGTATCGTGGACACACCGTGGGAAGGACACTTTGTTCTTAT 50
|||||
504 GlyThrValIleValAspSerThrValGlyLysAspThrLeuPheLeuII 520
51 CACCTGGACAAACGACCTCCCAATCTCTCTGGGATCCAGTGGAC 100
|||||
520 eThrTrpThrThrGlnProGlnIleLeuLeuTrpAspProSerGlyG 537

101	AGAACCAAGTGGCTTTGTTAGTGCACAAAACACACAAATGGCGCTACCTC	150
537	InLysGlnGlyGlyPheValValAspLysAsnThrLysMetAlaTyrLeu	553
151	CAAAATCCAGGCANTGCTAAGCTTGGCACTTGGAAATACAGCTGCAAGC	200
554	GlnIleProGlyIleAlaLysValGlyThrTrpLysTyrSerLeuGlnAl	570
201	AAGCTCACAACCTTGACCTGACTGTC	228
570	aSerSerGlnThrLeuThrVal	579
seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.AAB73716		
seq_documentation_block:		
ID	AAB73716 standard; Protein; 914 AA.	
XX		
AC	AAB73716;	
DT	11-SEP-2001 (first entry)	
XX		
DE	Human CLCA1 protein, SEQ ID NO:2.	
XX		
KW	Human CLCA1: goblet cell; mouse Gob-5 orthologue; drug screening;	
KW	expression inhibition; antisense therapy; gene therapy;	
KW	chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 152	
FT	/note= "Encoded by AGG in AAH46124"	
XX		
PN	W0200138530-Al.	
XX		
PD	31-MAY-2001.	
XX		
PF	22-NOV-2000; 2000WO-JP08232.	
XX		
PR	24-NOV-1999; 99JP-0333479.	
PR	27-APR-2000; 2000JP-0127589.	
XX		
PA	(TAKE) TAKEDA CHEM IND LTD.	
XX		
PI	Nakanishi A, Morita S;	
DR	WPI, 2001-355935/37.	
DR	N-PSDB; AAH46102, AAH46124.	
XX		
PT	New antisense nucleotide, useful for treatment and prevention of	
PT	bronchial asthma and chronic obstructive pulmonary disease	
XX		
PS	Claim 2; Page 76-80; 104pp; Japanese.	
XX		
CC	The invention relates to an antisense nucleotide targetted to the mouse	
CC	Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,	
CC	the CLCA1 gene (coding sequence shown in AAH46102). The invention also	
CC	relates to an antibody specific for the Gob-5 protein, medical and	
CC	diagnostic compositions containing the antisense nucleotide or the	
CC	antibody, and methods and kits for screening for compounds which inhibit	
CC	the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.	
CC	The antisense oligonucleotides and antibody are therefore useful for the	
CC	treatment and prevention of bronchial asthma and chronic obstructive	
CC	pulmonary disease. The present sequence represents human CLCA1 protein.	
XX		
SQ	Sequence	914 AA;
alignment_scores:		
	Quality: 396.00	Length: 76
	Ratio: 5.211	Gaps: 0
	Percent Similarity: 100.000	Percent Identity: 100.000

alignment_block:
US-09-049-696-10 x AAB73716

Align seg 1/1 to: AAB73716 from: 1 to: 914

1 GGCACAGTGCATCGTGCACACACCGTGGGAAAGGACACTTGTTCATTAT 50
|||||
504 GlyThrValIleValAspSerThrValGlyLysAspThrLeuPheLeuI 520
|||||
51 CACCTGGACACGACGCTCCCAAAATCTTCTCTGGATCCCGACGGAC 100
|||||
520 aThrTrpThrGlnProGlnIleLeuLeuTrpAspProSerGlyG 537
|||||
101 AGAAGCAAGTGGCTTTGTTAGTGCACAAAACACACAAATGGCGCTACCTC 150
|||||
537 InLysGlnGlyGlyPheValValAspLysAsnThrLysMetAlaTyrLeu 553
|||||
151 CAAATCCAGGCANTGCTAAGCTTGGCACTTGGAAATACAGCTGCAAGC 200
|||||
554 GlnIleProGlyIleAlaLysValGlyThrTrpLysTyrSerLeuGlnAl 570
|||||
201 AAGCTCACAACCTTGACCTGACTGTC 228
|||||
570 aSerSerGlnThrLeuThrVal 579
|||||
seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.AAB74733
seq_documentation_block:
ID AAB74733 standard; Protein; 914 AA.
XX
AC AAB74733;
XX
DT 12-JUN-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.
XX
KW Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
KW antialzheimer; antiparkinsonian; antimicrobial; vulnery; gene therapy;
KW immune disorder; hyperproliferative disorder; cardiovascular disease;
KW cancer; angiogenic disorder; neurological disorder; infectious disease;
KW wound healing; regeneration; chemotaxis; chromosome 1.
XX
OS Homo sapiens.
XX
PN W0200112775-A2.
XX
PD 22-FEB-2001.
XX
PF 16-AUG-2000; 2000WO-US22325.
XX
PR 17-AUG-1999; 99US-0149182.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, NI J, Florence KA, Fiscella M, Wei P, Baker KP;
PI Birse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;
XX
DR WPI, 2001-147550/15.
DR N-PSDB; AAF81787.
XX
PT Nucleic acids, encoding 25 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
PS Claim 11; Page 459-460; 485pp; English.
XX
CC AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733
CC to AAB74772. Human secreted proteins can have activities based on the
CC tissues and cells they are expressed in. Example of activities include:
XX

CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;
CC antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;
CC vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic;
CC anticonvulsant; antialzheimers; antiparkinsonian; antimicrobial; and
CC vulnerary. Human secreted proteins can be used in gene therapy and
CC vaccine. Human secreted protein nucleotide sequences (NAM1) and proteins
CC (PEP1) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. For example, NAM1
CC and PEP1 may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patients genome
CC that affect the activity of proteins by expressing inactive proteins or
CC to supplement the patients own production of polypeptides. Disorders that
CC may be prevented, diagnosed and/or treated include immune disorders,
CC hyperproliferative disorders (e.g. cancers), cardiovascular diseases,
CC angiogenic disorders, neurological disorders, infectious diseases and/or
CC for promoting wound healing, regeneration and /or chemotaxis. AAF81778 to
CC AAF81786 and AAB74732 represent sequences used in the exemplification of
CC the present invention.

XX SQ Sequence 914 AA;

alignment_scores:
Quality: 396.00 Length: 76
Ratio: 5.211 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-10 x AAB74733 ..

Align seq 1/1 to: AAB74733 from: 1 to: 914

1 GGCACAGTGCATCTGGACAGCACCGTGGGAAAGGACACTTTGTTCTTAT 50
|||||
504 GlyThrValIleValAspSerThrValGlyLysAspThrLeuPheLeuI 520
51 CACCTGGACAAACCGCTCCCAATCCCTCTCGGATCCCGACGTGGAC 100
|||||
520 eThrTrpThrGlnProGlnIleLeuLeuTrpAspProSerGlyG 537
101 AGAAGCAAGGTGGCTTTGTAGTGGACAAACACCAAAATGGCTACCTC 150
|||||
537 InLysGlnGlyGlyPheValValAspLysAsnThrLysMetAlaTyrLeu 553
151 CAAATCCAGGCATTCGTAGGTGGCACTTGGAAATACAGTCTGCAAGC 200
|||||
554 GlnIleProGlyIleAlaLysValGlyThrTrpLysTy+SerLeuGlnAl 570
201 AGCTCACAACCTTGACCTGACTGTC 228
|||||
570 aSerSerGlnThrLeuThrLeuThrVal 579

seq_name: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT: AAG75474

seq_documentation_block:

ID AAG75474 standard; Protein: 925 AA.

XX AC AAG75474;

XX DT 03-SEP-2001 (first entry)

XX DE Human colon cancer antigen protein SEQ ID NO:6238.

XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; chromosome 1.

XX OS Homo sapiens.

XX PN WO200122920-A2.

XX PD 05-APR-2001.

XX PF 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPT; 2001-235357/24.
DR N-PSDB: AAB34879.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX claim 11; Page 7686-7690; 9803pp; English.
XX
CC AAB32943 to AAB37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAB37196 to AAB37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 925 AA;

alignment_scores:

Quality: 396.00 Length: 76
Ratio: 5.211 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-10 x AAG75474 ..

Align seq 1/1 to: AAG75474 from: 1 to: 925

1 GGCACAGTGCATCTGGACAGCACCGTGGGAAAGGACACTTTGTTCTTAT 50
|||||
515 GlyThrValIleValAspSerThrValGlyLysAspThrLeuPheLeuI 531
51 CACCTGGACAAACCGCTCCCAATCCCTCTCGGATCCCGACGTGGAC 100
|||||
531 eThrTrpThrGlnProGlnIleLeuLeuTrpAspProSerGlyG 548
101 AGAAGCAAGGTGGCTTTGTAGTGGACAAACACCAAAATGGCTACCTC 150
|||||
548 InLysGlnGlyGlyPheValValAspLysAsnThrLysMetAlaTyrLeu 564
151 CAAATCCAGGCATTCGTAGGTGGCACTTGGAAATACAGTCTGCAAGC 200
|||||
565 GlnIleProGlyIleAlaLysValGlyThrTrpLysTy+SerLeuGlnAl 581
201 AGCTCACAACCTTGACCTGACTGTC 228
|||||
581 aSerSerGlnThrLeuThrLeuThrVal 590

seq_name: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74822

seq_documentation_block:

ID AAB74822 standard; Protein: 913 AA.

```

XX AC AAB74822;
XX DT 13-JUN-2001 (first entry)
XX DE Murine ICACC-1 protein sequence.
XX KW ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW inflammatory bowel disease; autoimmune disease.
XX OS Mus sp.
XX PN WO9944620-A1.
XX PD 10-SEP-1999.
XX PF 03-MAR-1999; 99WO-US04703.
XX PR 03-MAR-1998; 98US-0076815.
XX PA (MAGA-) MAGAININ PHARM INC.
XX PI Holroyd KT, Levitt RC, Maloy WL, Louahed J, McLane M;
XX PI Nicolaides NC, Zhou Y, Dong Q;
XX DR WPI; 1999-550979/46.
XX DR N-PSDB; AAF81925.
XX PT New nucleic acid encoding calcium activated chloride channel, used to
XX PT identify, e.g. specific modulators for treating atopic allergy -
XX PS Claim 12; Fig 2; 75pp; English.
XX CC The present sequence represents the murine interleukin 9 (IL-9) induced
XX CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
XX CC have anti-allergic, anti-asthmatic, anti-inflammatory and
XX CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
XX CC used to alleviate asthma (or more generally atopic allergy), while those
XX CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
XX CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
XX CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
XX CC (increased levels) or IBD (reduced levels), also for monitoring
XX CC treatment of these conditions. The ICACC proteins can be used:
XX CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
XX CC reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
XX CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
XX CC identify modulators and binding partners. ICACC polynucleotides can be
XX CC used to generate transgenic animals or recombinant cells, used to screen
XX CC for antagonists, also as a source of therapeutic antisense agents or
XX CC diagnostic probes (for quantifying mRNA expression, e.g. for
XX CC identification of modulators).
XX SQ Sequence 913 AA;

alignment_scores:
Quality: 319.00 Length: 76
Ratio: 4.493 Gaps: 0
Percent Similarity: 93.421 Percent Identity: 76.316

alignment_block:
US-09-049-696-10 x AAB74822 ..
Align seg 1/1 to: AAB74822 from: 1 to: 913
1 GGACAGTGTCTGGACAGCAGCGTGGGAAGACACTTGTCTTAT 50
505 GlySerValIleValAspSerValIleGlyLysAspThrLeuPheLeuI 521
51 CACCTGGACAAACGAGCGCTCCCAAACTCTCTCTGGGATCCAGTGAGC 100

```

```

521 eThrTrpThrHisProThrIlePheIleTrpAspProSerGlyV 538
101 AGAAGCAAGTGGCTTTGACTGGACAAAACACCAAAATGCGCTACCTC 150
538 aIcIuGlnAsnGlyPheIleLeuAspThrThrLysValAlaIleLeu 554
151 CAAATCCCAGGCAATGCTAAGGTTGGCACTTGGAAATACAGACTCTGCAAGC 200
555 GlnValProGlyThrAlaLysValGlyPheTrpLysTyrSerIleGlnAl 571
201 AGCTCACAAACCTTGACCTGACTGTC 228
571 aSerSerGlnThrLeuThrVal 580

seq_name: ./SIDS2/jcgdata/geneseq/geneseq/AA2001.DAT: AAB73715
seq_documentation_block:
ID AAB73715 standard; Protein; 913 AA.
XX AC AAB73715;
XX DT 11-SEP-2001 (first entry)
XX DE Mouse Gob-5 protein, SEQ ID NO:1.
XX KW Mouse Gob-5; murine; goblet cell; human CLCA1 orthologue; drug screening;
KW expression inhibition; antisense therapy; gene therapy; bronchial asthma;
KW chronic obstructive pulmonary disease; antiasthmatic.
XX OS Mus sp.
XX PN WO200138530-A1.
XX PD 31-MAY-2001.
XX PF 22-NOV-2000; 2000WO-JP08232.
XX PR 24-NOV-1999; 99JP-0333479.
XX PR 27-APR-2000; 2000JP-0127589.
XX PA (TAKÉ ) TAKEDA CHEM IND LTD.
XX PI Nakahishi A, Morita S;
XX DR WPI; 2001-355935/37.
XX DR N-PSDB; AAH46101, AAH46120.
XX PT New antisense nucleotide, useful for treatment and prevention of
XX PT bronchial asthma and chronic obstructive pulmonary disease -
XX PS Claim 1; Page 72-76; 104pp; Japanese.
XX CC The invention relates to an antisense nucleotide targetted to the mouse
XX CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
XX CC the CLCA1 gene (coding sequence shown in AAH46102). The invention also
XX CC relates to an antibody specific for the Gob-5 protein, medical and
XX CC diagnostic compositions containing the antisense nucleotide or the
XX CC antibody, and methods and kits for screening for compounds which inhibit
XX CC the protein.. Gob-5 and CLCA1 are proteins expressed by goblet cells.
XX CC The antisense oligonucleotides and antibody are therefore useful for the
XX CC treatment and prevention of bronchial asthma and chronic obstructive
XX CC pulmonary disease. The present sequence represents mouse Gob-5 protein.
XX SQ Sequence 913 AA;

alignment_scores:
Quality: 319.00 Length: 76
Ratio: 4.493 Gaps: 0
Percent Similarity: 93.421 Percent Identity: 76.316

alignment_block:

```


This Page Blank (uspto)

OM of: US-09-049-696-9 to: SPTREMBL_17.* out_format : pfs
Date: Mar 30, 2002 2:46 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+o2p.model -DEV=xlp
-O/cgcn2_1/USPRO.spool/US09049696/runat_28032002_145238_2053/app_query.fasta_1.12579
-DB-SPTREMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 @CGN1_1_705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-9
Query length: 218
Database: SPTREMBL_17.*
Database sequences: 473505
Database length: 146272329
Search time (sec): 805.760000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
sp_human:O95151	+	374.00	827.07	914	O95151 homo sapiens (human). ca
sp_human:Q9UPC6	+	374.00	827.07	914	Q9UPC6 homo sapiens (human). ca
sp_human:Q9UNF6	+	374.00	827.07	914	Q9UNF6 homo sapiens (human). ca
sp_rdent:O88826	+	313.00	688.08	913	O88826 mus musculus (mouse). g
sp_rdent:Q9D726	+	313.00	688.08	913	Q9D726 mus musculus (mouse). ch
sp_human:Q9NXP1	+	291.00	643.30	913	Q9NXP1 homo sapiens (human). co
sp_mammal:Q9UNT5	+	291.00	637.91	917	Q9UNT5 homo sapiens (human). ca
sp_rdent:Q9R074	+	220.00	476.24	902	Q9R074 mus musculus (mouse). ca
sp_rdent:Q9E0R4	+	215.00	464.85	902	Q9E0R4 mus musculus (mouse). en
sp_rdent:O88860	+	212.00	458.02	901	O88860 mus musculus (mouse). ch
sp_rdent:Q9QX15	+	212.00	458.01	902	Q9QX15 mus musculus (mouse). ca
sp_mammal:O18742	+	202.00	436.25	917	O18742 bos taurus (bovine). lu
sp_mammal:O18743	+	202.00	435.99	820	O18743 bos taurus (bovine). lu
sp_mammal:O18741	+	202.00	435.20	905	O18741 bos taurus (bovine). lu
sp_human:Q9UQC9	+	187.50	401.82	943	Q9UQC9 homo sapiens (human). ch
sp_human:Q9V6N2	+	187.50	401.82	943	Q9V6N2 homo sapiens (human). ca
sp_human:Q9BYN1	+	74.50	157.93	0.9191	Q9BYN1 homo sapiens (human). ke
sp_human:Q9BYN2	+	72.50	153.70	1.65	Q9BYN2 homo sapiens (human). ke
sp_human:Q9B762	+	72.50	153.32	1.65	Q9B762 homo sapiens (human). ke
sp_human:Q9BYE3	+	71.50	155.22	2.16	Q9BYE3 pseudoonas putida. out
sp_bacteria:Q9L950	+	69.00	125.16	4.98	Q9L950 pseudoonas putida. out
sp_vertebrate:Q9PUM2	+	68.00	127.60	6.54	Q9PUM2 xenopus laevis (african
sp_bacteria:Q9F0M3	+	67.00	135.03	8.43	Q9F0M3 enterococcus faecium (st
sp_bacteria:Q47755	+	67.00	135.03	8.43	Q47755 enterococcus faecium (st
sp_bacteria:Q9J2V5	+	67.00	129.55	8.61	Q9J2V5 chlamydia pneumoniae (ch
sp_rdent:Q9J122	+	67.00	123.85	8.82	Q9J122 mus musculus (mouse). r
sp_rdent:Q9J122	+	67.00	123.85	8.82	Q9J122 mus musculus (mouse). r
sp_rdent:Q9J122	+	67.00	123.85	8.82	Q9J122 mus musculus (mouse). r
sp_rdent:Q9J121	+	67.00	123.80	8.82	Q9J121 mus musculus (mouse). r
sp_rdent:Q9M81	+	67.00	123.79	8.82	Q9M81 mus musculus (mouse). r
sp_rdent:Q9M80	+	67.00	123.74	8.82	Q9M80 mus musculus (mouse). r
sp_human:Q36027	+	66.50	139.56	9.53	Q36027 homo sapiens (human). ke
sp_fungi:Q36028	+	66.50	123.10	10.19	Q36028 schizosaccharomyces pom
sp_rdent:Q9QX11	+	66.00	129.02	11.46	Q9QX11 mus musculus (mouse). co
sp_invertebrate:Q9W117	+	66.00	123.74	11.71	Q9W117 drosophila melanogast
sp_invertebrate:Q97441	+	65.50	135.17	12.87	Q97441 giardia lamblia (giard
sp_invertebrate:Q97449	+	65.50	135.17	12.87	Q97449 giardia lamblia (giard
sp_invertebrate:Q9PVV3	+	65.50	135.17	12.87	Q9PVV3 giardia lamblia (giard
sp_bacteria:Q9K958	+	65.50	129.77	13.16	Q9K958 bacillus halodurans. udi

sp_bacteria:Q9WVC7 + 65.00 135.40 14.82 194 ! Q9WVC7 agrobacterium radioba
sp_virus:Q71857 + 65.00 134.96 14.84 205 ! Q71857 human immunodeficienc
sp_virus:Q71857 + 65.00 134.96 14.84 205 ! Q71857 human immunodeficienc
sp_virus:Q71860 + 65.00 134.96 14.84 205 ! Q71860 human immunodeficienc
sp_virus:Q71861 + 65.00 134.96 14.84 205 ! Q71861 human immunodeficienc
seq_name: sp_human:O95151
seq_documentation_block:
ID O95151 PRELIMINARY; PRT; 914 AA.
AC O95151;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Elbie R.C., Ji H.L., Schreier K.D., Fuller C.M.,
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
of human CLCAL, the first human member of the family of Ca2+-activated
Cl-channel proteins.";
RL Genomics 54:200-214(1998).
DR EMBL; AF039400; AAC95428.1;
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS00234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 914 AA; 100174 MW; 7066440C46557FA3 CRC64;

alignment_scores:
Quality: 374.00 Length: 72
Ratio: 5.194 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-9 x O95151
Align seg 1/1 to: O95151 from: 1 to: 914
2 AACAAAGCCCTCATGATGCTTTGGGGCCCTTTCATCAGGAATGGAGC 51
|||||
465 AsnAsnGlyLeuLeuAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAl 481
52 TGCTCTCAGCGCTCCATCCAGCTTGAGTAAAGGATTAAACCCCTCCAGA 101
|||||
481 aValSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnA 498
102 ACAGCCAGTGGATGAATGGACACAGTATCGTGGACAGCAGCGTGGGAAG 151
498 snSerGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLys 514
152 GACACTTTGTTCTTATCCTCGACACAGCAGCGCTCCCAAAATCCCTT 201
515 AspThrLeuPheLeuIleThrTrpThrThrGlnProGlnIleLeuLe 531
202 CTGGGATCCCAAGTGA 217
531 utrpAspProSerGly 536

seq_name: sp_human:Q9UPC6

seq_documentation_block:
ID Q9UPC6 PRELIMINARY; PRT; 914 AA.

AC Q9UPC6;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
 GN HCLICAL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=99047526; PubMed=9828122;
 RX Gruber A.D., Elble R.C., Ji H.L., Schreur K.D., Fuller C.M.,
 RA Pauli B.U.;
 RT "Genomic cloning, molecular characterization, and functional analysis
 of human CLICAL, the first human member of the family of Ca2+-activated
 Cl-channel proteins.";
 RL Genomics 54:200-214(1998).
 DR EMBL; AF039401; AAC95429.1; -
 DR InterPro; IPR000131; ATPase_gamma.
 DR InterPro; IPR002035; VWFA.
 DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
 DR PROSITE; PS50234; VWFA; 1.
 DR SMART; SM00327; VWFA; 1.
 DR SEQUENCE 914 AA; 100160 MW; 7761362B66752AA3 CRC64;

alignment_scores:
 Quality: 374.00 Length: 72
 Ratio: 5.194 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-049-696-9 x Q9UPC6 ..
 Align seg 1/1 to: Q9UPC6 from: 1 to: 914

2 AACATGGCCTCATTTGATGCTTTTGGGGCCCTTTCATCAGGAATGGAGC 51
 |||||
 465 AsnAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAl 481
 52 TGCTCTCAGCGCTCCATCCAGCTTCAGAGTAAAGGATTAACCTCCAGA 101
 |||||
 481 aValSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnA 498
 |||||
 102 ACAGCCAGTGGATGAATGGCACAGTGCATGTCGACACGACCGCTGGGAAAG 151
 |||||
 498 snSerGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLys 514
 152 GACACTTTGTTCTTATCCTACCTGGACACGACGCTCCCAATCCTTCT 201
 |||||
 515 AspThrLeuPheLeuIleThrTrpThrGlnProGlnIleLeuLe 531
 202 CTGGGATCCCACTGGA 217
 |||||
 531 uTrpAspProSerGly 536

seq_name: sp_human:Q9UNF6

seq_documentation_block:
 ID Q9UNF6 PRELIMINARY; PRT; 914 AA.
 AC Q9UNF6;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 1.
 GN CACCL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE=SMALL INTESTINE, AND COLON;
 RX MEDLINE=99364503; PubMed=10437792;
 RA Agnel M., Vermet T., Culouscou J.M.;
 RT "Identification of three novel members of the calcium-dependent
 chloride channel (CaCC) family predominantly expressed in the
 digestive tract and trachea.";
 RL FEBS Lett. 455:295-301(1999).
 DR EMBL; AF127036; AAD25487.1; -
 DR InterPro; IPR000131; ATPase_gamma.
 DR InterPro; IPR002035; VWFA.
 DR SMART; SM00327; VWFA; 1.
 DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
 DR PROSITE; PS50234; VWFA; 1.
 DR SEQUENCE 914 AA; 100202 MW; B58B451D97887D6B CRC64;

alignment_scores:
 Quality: 374.00 Length: 72
 Ratio: 5.194 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-049-696-9 x Q9UNF6 ..
 Align seg 1/1 to: Q9UNF6 from: 1 to: 914

2 AACATGGCCTCATTTGATGCTTTTGGGGCCCTTTCATCAGGAATGGAGC 51
 |||||
 465 AsnAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAl 481
 52 TGCTCTCAGCGCTCCATCCAGCTTCAGAGTAAAGGATTAACCTCCAGA 101
 |||||
 481 aValSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnA 498
 |||||
 102 ACAGCCAGTGGATGAATGGCACAGTGCATGTCGACACGACCGCTGGGAAAG 151
 |||||
 498 snSerGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLys 514
 152 GACACTTTGTTCTTATCCTACCTGGACACGACGCTCCCAATCCTTCT 201
 |||||
 515 AspThrLeuPheLeuIleThrTrpThrGlnProGlnIleLeuLe 531
 202 CTGGGATCCCACTGGA 217
 |||||
 531 uTrpAspProSerGly 536

seq_name: sp_rodent:O88826

seq_documentation_block:
 ID O88826 PRELIMINARY; PRT; 913 AA.
 AC O88826;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE GCB-5 PROTEIN.
 GN CLCA3 OR GCB-5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=INTESTINE;
 RA Koniya T., Tanigawa Y., Hirohashi S.;
 RT "Cloning of a gene, gob-5, which is expressed in intestinal goblet
 cells in mice.";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB017156; BAA33743.1; -
 DR MGD; MGI:1346342; Clca3.
 DR InterPro; IPR000131; ATPase_gamma.
 DR InterPro; IPR002035; VWFA.
 DR SMART; SM00327; VWFA; 1.

DR PROSITE: PS00153; ATPASE_GAMMA; UNKNOWN_1.
 DR PROSITE: PS0234; VFPA; I.
 SQ SEQUENCE 913 AA; 100070 MW; A7FA2F9E1089806D CRC64;

alignment_scores:
 Quality: 313.00 Length: 72
 Ratio: 4.408 Gaps: 0
 Percent Similarity: 98.611 Percent Identity: 77.778

alignment_block:

US-09-049-696-9 x O88826 ..
 Align seg 1/1 to: O88826 from: 1 to: 913

2 AACATGGCCTCATGTGCTTTGGGCGCTTTCATCAGGAATGGAGC 51
 |||||
 466 AsnAsnGlyLeuValAspAlaPheAlaAlaLeuSerGlyAsnAlaA1 482
 52 TGCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAACCTCCAGA 101
 |||||
 482 aileAlaGlnHisSerIleGlnLeuGluSerArgGlyValAsnLeuGlnA 499
 |||||
 102 ACAGCCAGTGGATGAATGGCACAGCTGCTGGACAGCACCGTGGGAAAG 151
 |||||
 499 snAsnGlnTrpMetAsnGlySerValIleValAspSerValGlylys 515
 152 GACACTTGTGTTCTTATCAGCTGGACAGCAGCTCCCAATCCTTCT 201
 |||||
 516 AspThrLeuPheLeuIleThrTrpThrHisProThrIlePheI1 532
 202 CTGGATCCAGTCGA 217
 532 etrpAspProSerGly 537

seq_name: sp_rodent:Q9D726

seq_documentation_block:

ID Q9D726 PRELIMINARY; PRT; 913 AA.

AC Q9D726;
 DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE CHLORIDE CHANNEL CALCIUM ACTIVATED 3.
 GN CLCA3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=STOMACH;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant R.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kotsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).

DR EMBL; AK008659; BAB25815.1; ..
 DR MGD; MGI:1346342; Clca3.
 DR InterPro; IPR00131; ATPase_gamma.
 DR InterPro; IPR02035; VFPA.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
 DR PROSITE; PS0234; VFPA; I.
 SQ SEQUENCE 913 AA; 100111 MW; 491E8584B0609A89 CRC64;

alignment_scores:
 Quality: 313.00 Length: 72
 Ratio: 4.408 Gaps: 0
 Percent Similarity: 98.611 Percent Identity: 77.778

alignment_block:

US-09-049-696-9 x Q9D726 ..
 Align seg 1/1 to: Q9D726 from: 1 to: 913

2 AACATGGCCTCATGTGCTTTGGGCGCTTTCATCAGGAATGGAGC 51
 |||||
 466 AsnAsnGlyLeuValAspAlaPheAlaAlaLeuSerGlyAsnAlaA1 482
 52 TGCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAACCTCCAGA 101
 |||||
 482 aileAlaGlnHisSerIleGlnLeuGluSerArgGlyValAsnLeuGlnA 499
 |||||
 102 ACAGCCAGTGGATGAATGGCACAGCTGCTGGACAGCACCGTGGGAAAG 151
 |||||
 499 snAsnGlnTrpMetAsnGlySerValIleValAspSerValGlylys 515
 152 GACACTTGTGTTCTTATCAGCTGGACAGCAGCTCCCAATCCTTCT 201
 |||||
 516 AspThrLeuPheLeuIleThrTrpThrHisProThrIlePheI1 532
 202 CTGGATCCAGTCGA 217
 532 etrpAspProSerGly 537

seq_name: sp_human:Q9NXP1

seq_documentation_block:

ID Q9NXP1 PRELIMINARY; PRT; 469 AA.

AC Q9NXP1;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
 DE CDNA FLJ20131 FIS. CLONE COL06357.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COLON;
 RA Kiyabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
 RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK00138; BAA90969.1; ..
 SQ SEQUENCE 469 AA; 50694 MW; 6765986C886B3E56 CRC64;

alignment_scores:

Quality: 291.00 Length: 72
 Ratio: 4.409 Gaps: 0
 Percent Similarity: 91.667 Percent Identity: 75.000

alignment_block:

US-09-049-696-9 x Q9NXP1 ..

[illegible][illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

us-09-049-696-9 rspt

This Page Blank (uspto)

OM of: US-09-049-696-9 to: SwissProt_39.* out_format : pfs

Date: Mar 30, 2002 2:52 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+ntp.model -DEV=xlp
-O=/cgn2.1/USPTO_SPOOL/US09049696/funat_28032002.145238_2085/app_query.fasta_1.12579
-DB=SwissProt_39 -QFMT=fastn -SUFFIX=isp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPEXT=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPEXT=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DEPOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pcp
-THR_MAX=100 -THR_MIN=0 -ALICN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1_165 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-049-696-9
Query length: 218
Database: SwissProt_39.*
Database sequences: 100059
Database length: 36664827
Search time (sec): 306.030000
```

score_list:

Sequence	Strd Orig	Zscore	Escore	Len	Documentation
SwissProt_39:ECIC_BOVIN	+ 222.00	455.50	1.2e-18	903	! P54281 bos taurus (bovine).
SwissProt_39:ATIA_MOUSE	+ 75.50	137.45	0.4821	1149	! P70704 mus musculus (mouse).
SwissProt_39:ATIA_HUMAN	+ 75.50	137.34	0.4828	1164	! Q9Y260 homo sapiens (human).
SwissProt_39:ATIA_BOVIN	+ 68.50	122.35	3.34	1149	! Q29449 bos taurus (bovine).
SwissProt_39:RFA4_HUMAN	+ 65.50	128.68	6.53	261	! Q31586 homo sapiens (human).
SwissProt_39:CS9A_MOUSE	+ 64.50	133.03	7.94	123	! Q55186 mus musculus (mouse).
SwissProt_39:GUNS_BACAG	+ 64.50	122.84	9.02	400	! O85465 bacillus agaradhaerens.
SwissProt_39:DHFA_GITALA	+ 64.50	121.84	9.13	449	! P28724 giardia lamblia (glar).
SwissProt_39:YH2A_STRCO	+ 64.00	116.79	11.02	711	! Q33902 streptomyces coelicolor.
SwissProt_39:YH2A_YEAST	+ 63.50	121.67	11.75	357	! P38702 saccharomyces cerevisiae.
SwissProt_39:SM5B_MOUSE	+ 63.50	112.00	13.26	1093	! Q50519 mus musculus (mouse).
SwissProt_39:AMYG_YEAST	+ 62.50	115.79	16.23	549	! P08019 saccharomyces cerevisiae.
SwissProt_39:NUP1_YEAST	+ 62.50	109.98	17.45	1076	! P20676 saccharomyces cerevisiae.
SwissProt_39:PTB1_SCYLI	+ 62.00	123.26	16.75	204	! P15369 scytalidium lignicolum.
SwissProt_39:BID1_HAEIN	+ 62.00	121.79	17.06	242	! P45209 haemophilus influenzae.
SwissProt_39:EST1_CAEBP	+ 62.00	114.51	18.68	562	! Q04456 caenorhabditis briggsae.
SwissProt_39:ATCX_SCHPO	+ 62.00	106.62	20.62	1402	! Q09891 schizosaccharomyces.
SwissProt_39:IDMT_ECOLI	+ 61.50	115.57	20.89	439	! P39344 escherichia coli.
SwissProt_39:GRN_CAVPO	+ 61.50	113.00	21.57	591	! P28797 cavia porcellus (guinea pig).
SwissProt_39:AMVH_SACDI	+ 61.50	110.75	22.19	767	! P04065 saccharomyces diastaticus.
SwissProt_39:POLG_LANVT	+ 61.50	110.65	22.22	776	! P29838 langat virus (strain).
SwissProt_39:POLG_LANVT	+ 61.50	97.85	26.06	3414	! P29837 genome polyprotein.
SwissProt_39:MR55_YEAST	+ 60.50	125.44	23.71	109	! P32830 saccharomyces cerevisiae.
SwissProt_39:KR2A_SHEEP	+ 60.50	122.35	24.64	156	! P02439 ovis aries (sheep).
SwissProt_39:KR2D_SHEEP	+ 60.50	121.55	24.89	171	! P02438 ovis aries (sheep).
SwissProt_39:DNAA_PSEPU	+ 60.50	121.06	25.04	181	! P08131 ovis aries (sheep).
SwissProt_39:YAB4_MYCPN	+ 60.50	112.18	27.98	506	! P13454 pseudomonas putida.
SwissProt_39:YH32_STRCO	+ 60.50	111.88	28.08	524	! P75609 mycoplasma pneumoniae.
SwissProt_39:VDMB_BPT3	+ 60.00	115.25	30.50	313	! Q9K400 streptomyces coelicolor.
SwissProt_39:GRN_RAT	+ 60.00	109.84	32.84	586	! P10310 bacteriophage t3.
SwissProt_39:V120_HSV11	+ 60.00	109.81	32.65	588	! P23785 r granulin precursor (a).
SwissProt_39:ATIB_MOUSE	+ 60.00	104.22	35.01	1123	! P10221 herpes simplex virus.
SwissProt_39:ATC4_YEAST	+ 60.00	104.03	35.09	1148	! P98200 mus musculus (mouse).
SwissProt_39:AMV1_SACDI	+ 59.50	101.10	36.40	1612	! Q12675 saccharomyces cerevisiae.
SwissProt_39:AMV1_SACDI	+ 59.50	106.42	38.59	768	! P29760 saccharomyces diastaticus.
SwissProt_39:DRN2_HUMAN	+ 59.00	111.89	40.84	360	! Q00115 homo sapiens (human).
SwissProt_39:CATA_MYCSM	+ 59.00	105.68	44.13	739	! Q59557 mycobacterium smegmatis.
SwissProt_39:PIP1_DROME	+ 59.00	100.72	46.94	1312	! P25455 drosophila melanogaster.
SwissProt_39:VS48_TBRYV	+ 58.50	109.40	47.73	424	! P22051 tomato black ring virus.
SwissProt_39:SM6B_HUMAN	+ 58.50	103.01	51.69	888	! Q9H3T3 homo sapiens (human).

```
SwissProt_39:DDL_ENTHR + 58.00 109.78 53.82 358 ! Q47827 enterococcus hirae.
SwissProt_39:AMPC_ENTCL + 58.00 109.24 54.18 381 ! P05364 enterobacter cloac
SwissProt_39:VS48_TBRYV + 58.00 108.42 54.74 419 ! P22049 tomato black ring
SwissProt_39:Y4RF_RHISN + 58.00 108.30 54.82 425 ! P55639 rhizobium sp. (str
SwissProt_39:ACM2_PIG + 58.00 107.50 55.37 466 ! P06199 sus scrofa (pig). mu
```

seq_name: SwissProt_39:ECIC_BOVIN

seq_documentation_block:

```
ID ECIC_BOVIN STANDARD; PRT; 903 AA.
AC P54281;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN (CALCIUM-ACTIVATED CHLORIDE
DE CHANNEL).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Trachea;
RX MEDLINE=96125078; PubMed=8537359;
RA Cunningham S.A., Awayda M.S., Buben J.K., Ismailov I.I.,
RA Arrate M.P., Berdiev B.K., Benos D.J., Fuller C.M.;
RT "Cloning of an epithelial chloride channel from bovine trachea.";
RL J. Biol. Chem. 270:31016-31026(1995).
CC !- FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE
CC SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME;
CC MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND
CC TRANSEPITHELIAL TRANSPORT.
CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC !- TISSUE SPECIFICITY: TRACHEA.
CC !- PTM: PHOSPHORYLATED BY CAM-KINASE II.
CC !- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
EMBL: U36445; AAC48511.1; -
InterPro: IPR002035; VWFA.
SMART: SM00327; VWFA; 1.
PROSITE: PS50234; VWFA; 1.
Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
Calcium channel; Phosphorylation; Glycoprotein.
TRANSMEM 7 27 POTENTIAL.
TRANSMEM 331 351 POTENTIAL.
TRANSMEM 617 637 POTENTIAL.
TRANSMEM 883 903 POTENTIAL.
DOMAIN 308 476 VWFA.
FT CARBOHYD 75 75 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 688 688 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 811 811 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 816 816 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 842 842 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 857 857 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 903 AA; 100305 MW; 9742AB1590908AEC CRC64;
```

alignment_scores:

Quality: 222.00 Length: 71

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=99216279; PubMed=10198212;
 RA Moura I., Halleck M.S., Schlegel R.A., Mattei M.-G., Williamson P.L.,
 zachowski A., Devaux P., Cartton J.-P., Colin Y.;
 RT "Cloning, expression, and chromosomal mapping of a human ATPase II
 gene, member of the third subfamily of P-type ATPases and orthologous
 to the presumed bovine and murine aminophospholipid translocase.";
 RL Biochem. Biophys. Res. Commun. 257:333-339(1999).
 RN [2]
 RP SEQUENCE OF 4-1164 FROM N.A. (LONG ISOFORM).
 RC TISSUE=Brain;
 RA Osada S., Nakanishi Y.;
 RT "cDNA cloning of human ATPaseII.";
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE TRANSPORT OF AMINOPHOSPHOLIPIDS
 FROM THE OUTER TO THE INNER LEAFLET OF VARIOUS MEMBRANES AND THE
 MAINTENANCE OF ASYMMETRIC DISTRIBUTION OF PHOSPHOLIPIDS, MAINLY IN
 SECRETORY VESICLES.
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CHROMAFFIN
 GRANULES (BY SIMILARITY).
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG ISOFORM (SHOWN HERE)
 AND A SHORT ISOFORM; ARE PRODUCED BY TISSUE SPECIFIC ALTERNATIVE
 SPLICING.
 CC -!- TISSUE SPECIFICITY: FOUND IN MOST ADULT TISSUES EXCEPT LIVER,
 TESTIS AND PLACENTA. MOST ABUNDANT IN HEART, BRAIN AND SKELETAL
 MUSCLE. ALSO DETECTED IN FETAL TISSUES. THE LONG ISOFORM IS ONLY
 DETECTED IN BRAIN, SKELETAL MUSCLE AND HEART AND IS THE MOST
 ABUNDANT FORM IN SKELETAL MUSCLE.
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2
 ATPASES). SUBFAMILY IV.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF067820; AAD34706.1; -;
 DR EMBL; AB013452; BAA7248.1; -;
 DR InterPro; IPR001757; E1-E2_ATPase.
 DR Pfam; PF001454; Hydrolase.
 DR Pfam; PF00122; E1-E2_ATPase; 1.
 DR Pfam; PF00702; Hydrolase; 1.
 DR PRINTS; PR00119; CATATPASE.
 DR PROSITE; PS00154; ATPASE_E1_E2; 1.
 KW Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
 KW Alternative splicing; Multigene family.
 FT DOMAIN 1 65 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 66 86
 FT DOMAIN 87 92 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 93 115
 FT DOMAIN 116 297 POTENTIAL.
 FT TRANSMEM 298 319 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 320 344 POTENTIAL.
 FT TRANSMEM 345 366 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 367 857 POTENTIAL.
 FT TRANSMEM 858 878 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 879 890 POTENTIAL.
 FT TRANSMEM 891 910 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 911 940 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 941 962 POTENTIAL.
 FT DOMAIN 963 976 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 977 999 POTENTIAL.
 FT DOMAIN 1000 1005 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1006 1026 POTENTIAL.
 FT DOMAIN 1027 1044 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1045 1070 POTENTIAL.

FT DOMAIN 1071 1164 CYTOPLASMIC (POTENTIAL).
 FT MOD_RES 409 409 PHOSPHORYLATION (BY SIMILARITY).
 FT METAL 801 801 MAGNESIUM (BY SIMILARITY).
 FT METAL 805 805 MAGNESIUM (BY SIMILARITY).
 FT NP_BIND 741 748 ATP (POTENTIAL).
 FT NP_BIND 1095 1102 ATP (POTENTIAL).
 FT VARSPPLIC 433 447 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 364 364 E -> K (IN REF. 2).
 SQ SEQUENCE 1164 AA; 131369 MW; CELEAF0206CD36F7 CRC64;
 alignment scores:
 Quality: 75.50 Length: 64
 Ratio: 1.936 Gaps: 3
 Percent Similarity: 60.938 Percent Identity: 34.375
 alignment block:
 US-09-049-696-9 x ATIA_HUMAN
 Align_Seg 1/1 to: ATIA_HUMAN from: 1 to: 1164
 12 AACATGGCCTCATGTCGTTTGGGCGCTTCA...TCAGGAATGG 48
 225 AsnArgHisLeuTyrAspPheValGlyAsnIleArgLeuAspGlyHisG1 241
 49 AGCTGTCTCTCAGCGCTCCATCCAGCTGAGAGTAAGGATTACCCCTCC 98
 241 yThrValProLeuGlyAlaAspGlnIleLeuLeuArgGlyAlaGlnLeuA 258
 99 AGACAGCCAGTCGATGATGGCAGCAGTCGTCGGACACGACCGTGGGA 148
 258 rgnsnTrGlnTrpValHisGlyIleValValTyr.....ThrGly 271
 149 AAGGACACTTGTTCCTTATCACTGTCGACGACGACGCTCC 190
 272 HisAspThrLys...LeuMetGlnAsnSerThrSerProPro 284

seq_name: SwissProt_39:ATIA_BOVIN

seq_documentation_block:

ID ATIA_BOVIN STANDARD; PRT: 1149 AA.

AC Q29449;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE POTENTIAL PHOSPHOLIPID-TRANSPORTING ATPASE IA (EC 3.6.3.13)
 DE (CHROMAFFIN GRANULE ATPASE II).
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 502-678 AND 1144-1149.
 RC TISSUE=Adrenal medulla;
 RX MEDLINE=96234315; PubMed=8633245;
 RA Tang X., Halleck M.S., Schlegel R.A., Williamson P.L.;
 RT "A subfamily of P-type ATPases with aminophospholipid transporting
 activity.";
 RL Science 272:1495-1497(1996).
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE TRANSPORT OF AMINOPHOSPHOLIPIDS
 FROM THE OUTER TO THE INNER LEAFLET OF VARIOUS MEMBRANES AND THE
 MAINTENANCE OF ASYMMETRIC DISTRIBUTION OF PHOSPHOLIPIDS, MAINLY IN
 SECRETORY VESICLES.
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CHROMAFFIN
 GRANULES.
 CC -!- TISSUE SPECIFICITY: KIDNEY.
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2
 ATPASES). SUBFAMILY IV.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -


```

RA STRAIN=DSM 8721;
RA Bjoernvad M.E.;
RA Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.57 ANGSTROMS) OF 30-329.
RC STRAIN=AC13 / NCIMB 40482;
RX MEDLINE=98153671; PubMed=9485319;
RA Davies G.J., Dauter M., Brzozowski A.M., Bjoernvad M.E.,
RA Andersen K.V., Schueler M.;
RT "Structure of the Bacillus agaradherans family 5 endoglucanase at 1.6-
RT A and its cellobiose complex at 2.0-A resolution.";
RL Biochemistry 37:1926-1932(1998).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.64 ANGSTROMS) OF 30-329.
RC STRAIN=AC13 / NCIMB 40482;
RX MEDLINE=98384136; PubMed=9718293;
RA Davies G.J., Mackenzie L.F., Varrot A., Dauter M., Brzozowski A.M.,
RA Schueler M., Withers S.G.;
RT "Snapshots along an enzymatic reaction coordinate: analysis of a
RT retaining beta-glycoside hydrolase.";
RL Biochemistry 37:11707-11713(1998).
CC -|- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -|- SUBUNIT: MONOMER.
CC -|- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF067428; AAC19169.1; -
DR PDB; 1A3H; 16-MAR-99.
DR PDB; 2A3H; 16-MAR-99.
DR PDB; 3A3H; 16-MAR-99.
DR InterPro: IPR003610; Chitin_bind3.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam; PF00150; cellulase; 1.
DR SMART; SM00495; ChtB03; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; 3D-structure.
KW ACT_SITE 165 165 PROTON DONOR.
FT ACT_SITE 254 254 NUCLEOPHILE.
SQ SEQUENCE 400 AA; 44702 MW; 3F9C66FB9BC36FFF CRC64;

alignment_scores:
  Quality: 64.50 Length: 47
  Ratio: 2.016 Gaps: 1
Percent Similarity: 68.085 Percent Identity: 31.915

alignment_block:
US-09-049-696-9 x GUN5_BACAG ..
Align seg 1/1 to: GUN5_BACAG from: 1 to: 400
2 AACAAATGGCTCATTCATTCCTTTTGGGGCCCTTCATCAGGAATGGA... 49
28 AsnAspSerValValGluGluHisGlyGlnLeuSerIleSerAsnGlyG1 44
50 .GCTGTCCTCAGCGCTCCATCCACTGAGTAGAGTAGGAGTAACCCCTCC 98
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
44 uLeuValAsnGluArgGlyGluGlnValGlnLeuLysGlyMetSerSerH 61
99 AGNACAGCAGTAGGATGATGCACAGATGATCGTGACAGC 139
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 isGlyLeuGlnIlePyrGlyGlnPheValAsnTyrGluSer 74
seq_name: SwissProt_39:DHE4_GIALA

seq document_block:
ID DHE4_GIALA STANDARD: PRT: 449 AA.
AC P28724; Q24961;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE NADP-SPECIFIC GLUTAMATE DEHYDROGENASE (EC 1.4.1.4) (NADP-GDH).
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinales; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92218410; PubMed=1559991;
RA Yee J., Dennis P.P.;
RT "Isolation and characterization of a NADP-dependent glutamate
RT dehydrogenase gene from the primitive eucaryote Giardia lamblia.";
RL J. Biol. Chem. 267:7539-7544(1992).
RN [2]
RP SEQUENCE OF 59-261 FROM N.A.
RX MEDLINE=96155200; PubMed=8587793;
RA Monis P.T., Mayrhofer G., Andrews R.H., Homan W.L., Limper L.,
RA Ey P.L.;
RT "Molecular genetic analysis of Giardia intestinalis isolates at the
RT glutamate dehydrogenase locus.";
RL Parasitology 112:3-12(1996).
CC -|- CATALYTIC ACTIVITY: L-GLUTAMATE + H(2)O + NADP(+) = 2-OXOGLUTARATE
CC + NH(3) + NADPH.
CC -|- SUBUNIT: HOMOHEXAMER (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE GLU/LEU/PHE/VAL DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M84604; AAN29155.1; -
DR EMBL; U47632; AAB03400.1; -
DR PIR; A42489; A42489.
DR HSP; P24295; IAUP.
DR InterPro: IPR001625; GLFV_dehydrog.
DR Pfam; PF00208; GLFV_dehydrog; 1.
DR PRINTS; PR00082; GLFV_HDRGNASE.
DR PROSITE; PS00074; GLFV_DEHYDROGENASE; 1.
DR Oxidoreductase; NADP.
FT ACT_SITE 125 125 BY SIMILARITY.
FT CONFLICT 249 249 I -> L (IN REF. 2).
SQ SEQUENCE 449 AA; 49766 MW; 5497B35209B549F6 CRC64;

alignment_scores:
  Quality: 64.50 Length: 47
  Ratio: 2.016 Gaps: 2
Percent Similarity: 68.085 Percent Identity: 40.426

alignment_block:
US-09-049-696-9 x DHE4_GIALA ..
Align seg 1/1 to: DHE4_GIALA from: 1 to: 449
2 AACAAATGGCTCATTCATTCCTTTTGGGGCCCTTCATCAGGAATGAGC 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
223 AsnAsnThrValIleArgGlyLysAsnValLeuLeuSerGlySerGlyAs 239
52 TGTCCTCTCAG.....CGCTCCATCCAGCTTGAGAGTAAGGAT 89
|||||:|||||:|||||:|||||:|||||:|||||:
239 hValAlaGlnPheAlaCysGlnLysLeuIleGlnLeuGlyAlaLysVal 256
90 TAACCCCTCCAGAACAGCCAGTCGATGATGAGTGCACAGTCATC 130
|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: SwissProt_39:DHE4_GIALA
```

256 euThrPheSerAspSer.....AsnGlyThrIleVal 266

seq_name: SwissProt_39:MMLA_STRCO

seq_documentation_block:

ID MMLA_STRCO STANDARD; PRT; 711 AA.
AC Q53902;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PUTATIVE MEMBRANE PROTEIN ACTII-3.
GN ACTII-3.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91347376; PubMed=1878971;
RA Fernandez-Moreno M.A., Caballero J.L., Hopwood D.A., Malpartida F.;
RT "The act cluster contains regulatory and antibiotic export genes,
direct targets for translational control by the bldA tRNA gene of
Streptomyces";
RL Cell 66:769-780(1991).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE MMLP FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M64683; AAA26691.1;
DR InterPro; IPR000731; HMGCR_patched_5TM..
DR PROSITE; PS01556; SSD; 2.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 235 255 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 313 333 POTENTIAL.
FT TRANSMEM 369 389 POTENTIAL.
FT TRANSMEM 516 536 POTENTIAL.
FT TRANSMEM 540 560 POTENTIAL.
FT TRANSMEM 573 593 POTENTIAL.
FT TRANSMEM 623 643 POTENTIAL.
FT TRANSMEM 645 665 POTENTIAL.
SQ SEQUENCE 711 AA; 74862 MW; A5466BEDABED1B6 CRC64;

alignment_scores:
Quality: 64.00 Length: 43
Ratio: 2.207 Gaps: 0
Percent Similarity: 67.442 Percent Identity: 32.558

alignment_block:

US-09-049-696-9/rev x MMLA_STRCO ..

Align seg 1/1 to: MMLA_STRCO from: 1 to: 711

213 CTGGATCCAGAGCAAGCATTTGGGAGCGTGTTCAGGTGATAAG 164
|||||
301 LeuAsnSerThrLysGlyLeuGlyProValCysAlaValGlyValLeuVa 317
163 AAACAAGTGTCTTCCCGAGCGTCGTCCAGACACTGTCCCAATCA 114
|||
317 lGlyLeuLeuSerMetMetThrLeuLeuProAlaLeuLeuValIlePheG 334
113 TCCACTGGCTGTCTGGAGGGTTAATCCC 85

334 lyArgTrpValPheTrpProAlaArgPro 343

seq_name: SwissProt_39:YHG2_YEAST

seq_documentation_block:

ID YHG2_YEAST STANDARD; PRT; 357 AA.
AC F38702;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE PUTATIVE MITOCHONDRIAL CARRIER YHR002W.
GN YHR002W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du'Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kuzaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA LaReille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
VIII";
KL Science 265:2077-2082(1994).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC -!- INNER MEMBRANE (POTENTIAL).
CC -!- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U10555; AAB68424.1;
DR PIR; S46795; S46795.
DR SGD; S0001044; YHR002W.
DR InterPro; IPR002067; Mit_carrier.
DR InterPro; IPR001993; Mitoch_carrier.
DR Pfam; PF00153; mito_carr; 3.
DR PRINTS; PR00926; MITOCARRIER.
DR PROSITE; PS00215; MITOCH_CARRIER; 3.
KW Hypothetical protein; Mitochondrion; Inner membrane; Repeat;
KW Transmembrane; Transport.
FT TRANSMEM 136 153 POTENTIAL.
FT TRANSMEM 208 228 POTENTIAL.
FT TRANSMEM 325 347 POTENTIAL.
SQ SEQUENCE 357 AA; 40825 MW; 16266B2CD4B996D5 CRC64;

alignment_scores:
Quality: 63.50 Length: 70
Ratio: 1.549 Gaps: 3
Percent Similarity: 58.571 Percent Identity: 25.714

alignment_block:

US-09-049-696-9/rev x YHG2_YEAST ..

Align seg 1/1 to: YHG2_YEAST from: 1 to: 357

204 AGAAGGATTTGGGAGCGTGTTCAGGTGATAAGAAACAAGTCTC 152
|||
167 LysArgValLysLeuGlyArgIleIleLysLysIleTyrLysGluProAl 183

us-09-049-696-9-rsp

Tue Apr 2 09:40:16 2002

FT TRANSMEM 979 999
FT DOMAIN 1000 1093
FT DOMAIN 236 518
FT DOMAIN 551 605
FT DOMAIN 606 663
FT DOMAIN 664 720
FT DOMAIN 721 776
FT DOMAIN 795 851
FT DOMAIN 852 908
FT DOMAIN 909 958
FT CARBOHYD 59 59
FT CARBOHYD 95 95
FT CARBOHYD 157 157
FT CARBOHYD 178 178
FT CARBOHYD 287 287
FT CARBOHYD 333 333
FT CARBOHYD 378 378
FT CARBOHYD 532 532
FT CARBOHYD 539 539
FT CARBOHYD 547 547
FT CARBOHYD 602 602
FT CARBOHYD 728 728
FT CARBOHYD 944 944
SQ SEQUENCE 1093 AA; 120326 MW; 29E5C9B1EB108717 CRC64;

alignment_scores: Quality: 63.50 Length: 37
Ratio: 2.646 Gaps: 2
Percent Similarity: 64.865 Percent Identity: 40.541
alignment_block:
US-09-049-696-9/rev x SM5B_MOUSE ..
Align seg 1/1 to: SM5B_MOUSE from: 1 to: 1093
153 AAGTGTCTTCCACGGTGTCTCCAGCATCTGCATTCATCCAC 109
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
492 ArgileProLeuGluArgCys.....SerAlaTyHisSerGlnG1 505
108 TGGCTGTCTCGAGGTTAATCCCTTACTCTCAAGCTGGATGGAGCGCTG 59
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
505 yAlacysLeuGlyAlaArgAspProTyr.....CysGlyTriaSpGlyL 520
53 AGAGACAGCTC 48
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
520 ysArgGlnLeu 523
seq_name: SwissProt_39:SM5B_MOUSE

seq_documentation_block:
ID SM5B_MOUSE STANDARD; PRT; 1093 AA.
AC Q60519; 2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE SEMAPHORIN 5B PRECURSOR (SEMAPHORIN G) (SEMA G).
GN SEMA5B OR SEMAG OR SEMG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96414430; PubMed=8817451;
RA Adams R.H., Betz H., Puschel A.W.;
RT "A novel class of murine semaphorins with homology to thrombospondin
is differentially expressed during early embryogenesis.";
RL Mech. Dev. 57:33-45(1996).
CC -!- FUNCTION: MAY ACT AS POSITIVE AXONAL GUIDANCE CUES.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: IN ADULT, ONLY DETECTED IN BRAIN.
CC -!- DEVELOPMENTAL STAGE: DIFFERENTIALLY EXPRESSED IN EMBRYONIC AND
ADULT TISSUES. ITS ABUNDANCE DECREASES FROM E10 TO BIRTH.
CC -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-1 DOMAINS.

seq_documentation_block:
ID SM5B_MOUSE STANDARD; PRT; 1093 AA.
AC Q60519; 2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE SEMAPHORIN 5B PRECURSOR (SEMAPHORIN G) (SEMA G).
GN SEMA5B OR SEMAG OR SEMG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96414430; PubMed=8817451;
RA Adams R.H., Betz H., Puschel A.W.;
RT "A novel class of murine semaphorins with homology to thrombospondin
is differentially expressed during early embryogenesis.";
RL Mech. Dev. 57:33-45(1996).
CC -!- FUNCTION: MAY ACT AS POSITIVE AXONAL GUIDANCE CUES.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: IN ADULT, ONLY DETECTED IN BRAIN.
CC -!- DEVELOPMENTAL STAGE: DIFFERENTIALLY EXPRESSED IN EMBRYONIC AND
ADULT TISSUES. ITS ABUNDANCE DECREASES FROM E10 TO BIRTH.
CC -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-1 DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
EMBL; X97818; CAA66398.1; -
MGI; 107555; Sema5b.
MGI; 107555; Sema5b.
InterPro: IP0003659; PSI.
InterPro: IP0002165; Plexin_repeat.
InterPro: IP0001627; Sema.
InterPro: IP000884; TSP1.
Pfam: PF01437; Plexin_repeat; 1.
Pfam: PF01403; Sema; 1.
Pfam: PF00090; TSP-1; 5.
SMART; SM00423; PSI; 1.
SMART; SM00209; TSP1; 4.
PROSITE; PS50092; TSP1; 6.
Signal; Transmembrane; Repeat; Multigene family; Neurogenesis;
Developmental protein; Glycoprotein.
POTENTIAL.
SEMAPHORIN 5B.
CHAIN 20 1093
DOMAIN 20 978
EXTRACELLULAR (POTENTIAL).

seq_documentation_block:
ID AMYG_YEAST STANDARD; PRT; 549 AA.
AC P08019; 1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE GLUCOAMYLASE, INTRACELLULAR SPORULATION-SPECIFIC (EC 3.2.1.3) (GLUCAN
1,4-ALPHA-GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE).
DE SGAL OR SGA OR YIL099W
GN Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87194600; PubMed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
STAL.";
RL J. Bacteriol. 169:2142-2149(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;

```

01 NOV-1997 (Rel. 35, Last annotation update)
DE NUCLEOPORIN NUP1 (NUCLEAR PORE PROTEIN NUP1).
GN NUP1 OR YOR098C OR YOR3182C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
CC NCBI_TaxID=4932;
RN [1].
RN [2].
RP SEQUENCE FROM N.A.
RX MEDLINE=90275616; PubMed=2190694;
RA Davis L.I., Fink G.R.;
RT "The NUP1 gene encodes an essential component of the yeast nuclear
RT pore complex.";
RL Cell 61:965-978(1990).
RN [1].
RN [2].
RP SEQUENCE FROM N.A.
RX MEDLINE=97344368; PubMed=9200815;
FA Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C.,
RA Schwager C., Paces V., Sander C., Ansorge W.;
RT "RNA sequencing and analysis of 130 kb from yeast chromosome XV.";
RL yeast 13:655-672(1997).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
CC -1- NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
CC -1- THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.
CC -1- DOMAIN: APPEARS TO BE DIVIDED INTO THREE DOMAINS DEFINED BY
CC -1- CENTRALLY LOCATED REPEATING UNITS. FUNCTIONAL N-TERMINAL OF NUP1
CC -1- OR OF NUP2 IS REQUIRED FOR GROWTH.
CC -1- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
CC -1- SIMILARITY: TO THE CENTRAL REPEATING UNITS OF NSP1 AND NUP2, AND
CC -1- TO MAMMALIAN P62.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR ENBL; M33632; AAA34822.1; -
DR ENBL; X94335; CAA64020.1; -
DR ENBL; Z75006; CAA99295.1; -
DR PIR; A35622; A35622.
DR SGD; S0005624; NUP1.
KW Nuclear protein; Transport; Repeat.
FT DOMAIN 333 949
FT SEQUENCE 1076 AA; 113581 MW; 4AC23567D2FB53CC CRC64;
CC
CC alignment_scores:
CC Quality: 62.50 Length: 68
CC Ratio: 1.736 Gaps: 2
CC Percent Similarity: 52.941 Percent Identity: 30.882
CC
CC alignment_block:
CC US-09-049-696-9 x NUP1_YEAST
CC
CC Align_Seg 1/1 to: NUP1_YEAST from: 1 to: 1076
CC
CC 35 TCATCAGGAAATGGAGCTGTCTCTCAGCGCTCCATCCAGCTT..... 76
CC ::::||||| ::: :|||||:
CC 354 ThrSerAlaGlyAlaValPheLysSerValGluMetGlyLysTh 370
CC
CC 77 .....GAGATMAGGGATTACCCCTCCAGAAC 104
CC :||||| :|||||
CC 370 rAsPlySerThrLysThrAlaGluAlaProThrLeuSerPheAsnPheS 387
CC
CC 105 GCCATGGATCAATGGCAGCTGATCGTGGACAGCACCGTGGGAAGGAC 154
CC :||||| :||| :||| :|||||:||||| :::
CC 387 erGlnLysAlaAsnLysThrLysAlaValAspAsnThrValProSerThr 403
CC
CC 155 ACTTGTTT.....CTATCAGCTGCAGCAACGCA 183

```

us-09-049-696-9.rsp

Tue Apr 2 09:40:16 2002

```

|||||
404 ThrLeuPheAsnPhcGlyGlySerAspThrValThrSerAlaSerG1 420
184 GGCT 187
420 nPro 421

seq_name: SwissProt_39:PRTB_SCYLI
seq_documentation_block:
ID PRTB_SCYLI STANDARD; PRT; 204 AA.
AC P15369.
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE SCYTTALIDOPEPSIN B (EC 3.4.23.32) (ACID PROTEASE B) (SLB).
OS Scytalidium lignicolum.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Scytalidium.
OX NCBI_TaxID=5539;
RN [1]
RP MEDLINE=84185536; PubMed=6370989;
RX Maita T., Nagata S., Matsuda G., Maruta S., Oda K., Murao S.,
RA Tsuru D.;
RT "Complete amino acid sequence of Scytalidium lignicolum acid protease
B.";
RL J. Biochem. 95:465-475(1984).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS WITH BROAD SPECIFICITY,
CC CLEAVING 24-PHE-1-PHE-25, BUT NOT 15-LEU-TYR-16 AND 25-PHE-TYR-26
CC IN THE B CHAIN OF INSULIN.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A4.
CC PIR: A28864; A28864.
DR MEROPS: A04.001; -.
DR InterPro: IPR000250; Peptidase_A4.
DR Pfam: PF01828; Peptidase_A4; 1.
DR PRINTS: PR00977; SCYTTDPASE.
DR ProDom: PD018627; Peptidase_A4; 1.
KW Hydrolase; Aspartyl protease.
FT ACT_SITE 53
FT ACT_SITE 98
FT DISULFID 47
FT DISULFID 140
FT DISULFID 192
FT DISULFID 201
FT SEQUENCE 204 AA; 21362 MW; C03530AB74C86D34 CRC64;
SQ

```

alignment_scores:
Quality: 62.00 Length: 43
Ratio: 2.214 Gaps: 3
Percent Similarity: 65.116 Percent Identity: 41.860

alignment_block:

US-09-049-696-9/rev x PRTB_SCYLI

Align seg 1/1 to: PRTB_SCYLI from: 1 to: 204

```

192 TGGGAGGCTGGTGTCCAGGTGATAGAAACAAAGTCTCTTCCAC 143
|||||
6 TrpGlyGlyAlaLeu.....IleGlySerAspPheAspTh 18
142 GTGCTGTCCAGCATCTGTCATTCATCCACTGGCTGTCTGGAGG 93
|||||
18 rValSerAlaThrAlaAsnValPro.SerAlaThrGlyAlaSerGly 34
92 TTAATCCCTTACTCTCAAGCTGGATG 66
35 .....SerSerAlaAlaTrp 39

```

seq_name: SwissProt_39:BID1_HAEIN

seq_documentation_block:

ID BID1_HAEIN STANDARD; PRT; 242 AA.

```

AC P45209;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE DETHIOBIOTIN SYNTHETASE 1 (EC 6.3.3.3) (DETHIOBIOTIN SYNTHASE
DE 1) (DTB SYNTHETASE 1) (DTBS 1).
CN BIOD-A OR H11445.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney A., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd.";
RL Science 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: ATP + 7,8-DIAMINONONANOATE + CO(2) = ADP +
CC ORTHOPHOSPHATE + DETHIOBIOTIN.
CC -1- COFACTOR: REQUIRES MAGNESIUM (BY SIMILARITY).
CC -1- PATHWAY: BIOTIN BIOSYNTHESIS PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE DETHIOBIOTIN SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC ENEL; U32823; AAC23095.1;
CC HESP; P13000; IDTS.
CC TIGR; H11445; -.
CC Biotin biosynthesis; Ligase; Magnesium; ATP-binding;
CC Complete proteome.
CC NP_BIND 8 16 ATP (BY SIMILARITY).
CC SEQUENCE 242 AA; 26967 MW; 94FCC0501CFAF262 CRC64;
SQ

```

alignment_scores:
Quality: 62.00 Length: 41
Ratio: 2.480 Gaps: 1
Percent Similarity: 60.976 Percent Identity: 34.146

alignment_block:

US-09-049-696-9/rev x BID1_HAEIN

Align seg 1/1 to: BID1_HAEIN from: 1 to: 242

```

186 GGCTGCTGCTTCCAGGTGATAGAAACAAAGTCTCTTCCACGCTGCT 137
|||||
159 GlycysVal.....AsnHisAlaLeuThrVallysVal 171
136 GTCCACGATCACTGTCATTCATCCACTGGCTGTCTGGAGGTTAATC 87
|||||
171 eGlnGlnLeuGlyValProLeuLeuGlyTrpIleAlaAsnArgIleAsn 188
86 CCTTACTCTCAAGCTGGATGAG 64
183 roLeuLeuSerHistyrAlaGlu 195

```


OM of: US-09-049-696-9 to: PIR_68:* out_format : pfs

Date: Mar 30, 2002 2:26 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters: -DEV-xmlp

-MODE=frame-n2p.model -DBP=1.000 -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -GAPOP=4.500
-GAPEXT=0.050 -GAPOP=10.000 -GAPEXT=0.500 -GAPOP=6.000
-GAPEXT=7.000 -GAPOP=10.000 -GAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09049696 -ECGNI_1_367 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-049-696-9

Query length: 218

Database: PIR_68*

Database sequences: 219241

Database length: 76174552

Search time (sec): 470.790000

score_list:	Strd	Orig	ZScore	EScore	Len	Documentation
Sequence	+	313.00	659.08	1.1e-29	913	! gob-5 protein - mouse
Pir2:TJ00168	+	202.00	416.91	3.5e-16	905	! Lu-ECAM-1 protein - bovine
Pir2:TJ02205	+	72.50	139.00	0.8214	1149	! probable adenosinetriphosphata
Pir2:T30869	+	75.50	147.05	1.92	175	! high-sulfur keratin - human
Pir2:S37649	+	68.50	123.72	5.83	1149	! adenosinetriphosphatase (EC 3
Pir2:T818515	+	67.00	124.22	8.89	707	! oligopeptide binding lipoprotein
Pir2:D85656	+	67.00	124.22	8.89	707	! peptide ABC transporter, peripl
Pir2:E72059	+	66.50	133.87	10.29	177	! high-sulfur keratin - human
Pir2:T37650	+	66.50	133.87	10.29	177	! probable calcium-transporting
Pir2:T38820	+	65.50	118.01	10.19	1367	! UDP-N-acetylmuramoylalanine-D-9
Pir2:T83970	+	64.50	124.45	13.56	450	! glutamate dehydrogenase (NADP+)
Pir1:A42489	+	64.50	122.28	17.93	449	! MHC class I histocompatibility
Pir2:I50106	+	64.00	126.41	20.70	229	! hypothetical protein PA4959 [im
Pir2:E83027	+	64.00	117.85	20.58	691	! antibiotic transport-associated
Pir2:C40046	+	64.00	117.62	20.58	711	! hypothetical protein YHR002w -
Pir2:S46795	-	63.50	121.88	23.75	357	! latent transforming growth fac
Pir2:A55494	-	63.50	109.24	23.56	1820	! ferric vibriobactin ABC transpo
Pir2:A82281	+	63.00	121.52	27.34	325	! glucan 1,4-alpha-glucosidase (E
Pir1:S48474	+	62.50	116.36	31.36	549	! nuclear pore protein NUP1 - ye
Pir1:A35622	+	62.50	111.14	31.26	1076	! hypothetical protein KIAA0708
Pir2:T00350	-	62.50	107.35	31.18	1753	! proteinase B (EC 3.4.23.-) - fu
Pir2:A28864	-	62.00	122.95	36.25	204	! dethiobiotin synthetase homolog
Pir1:I64123	+	62.00	121.62	36.22	242	! esterase precursor - Caenorhab
Pir2:S27782	+	62.00	115.08	36.07	562	! probable calcium-transporting
Pir2:S62557	+	62.00	107.99	35.90	1402	! endo-1,4-beta-xylanase (EC 3.2.
Pir2:JC7307	+	61.50	121.23	41.68	221	! hypothetical 46.0K protein (pep
Pir2:S56491	+	61.50	115.91	41.53	439	! acrogranin - guinea pig (fragme
Pir2:I48141	-	61.50	113.60	41.47	591	! probable trp protein H (trph)
Pir2:G71302	-	61.50	112.37	41.44	693	! genome polyprotein - Langat vir
Pir2:A41704	+	61.50	111.49	41.42	776	! genome polyprotein - Langat vir
Pir2:B41704	+	61.50	111.49	41.42	776	! glucan 1,4-alpha-glucosidase (E
Pir1:ALBYG	+	61.50	111.47	41.42	778	! hypothetical protein TGH20.50
Pir2:T12956	+	61.00	108.30	41.33	1171	! hypothetical protein DKFZp564M2
Pir2:T12456	+	61.00	115.80	47.80	387	! F23A5.27 [imported] - Arabidops
Pir2:A96842	+	61.00	113.61	47.74	513	! MRS5 protein - yeast (Saccharom
Pir2:S30793	-	60.50	124.54	55.33	109	! high-sulfur wool matrix protein
Pir2:I47111	-	60.50	121.96	55.24	152	! high-sulfur wool matrix protein
Pir2:I47112	-	60.50	121.96	55.24	152	! high-sulfur wool matrix protein
Pir2:I47108	-	60.50	121.96	55.24	152	! keratin high-sulfur matrix prot
Pir1:KRSHHB	-	60.50	121.75	55.23	156	! high-sulfur wool matrix protein
Pir2:I47107	-	60.50	121.46	55.22	162	! high-sulfur wool matrix protein
Pir1:KRSHHA	-	60.50	121.00	55.20	172	! keratin high-sulfur matrix prot

Pir2:I47106 - 60.50 121.00 55.20 172 ! high-sulfur wool matrix prot
Pir1:KRSHHD - 60.50 120.56 55.19 182 ! keratin high-sulfur matrix p
Pir2:I47105 - 60.50 120.56 55.19 182 ! high-sulfur wool matrix prot
Pir2:SS5621 - 60.50 119.94 55.17 197 ! hypothetical protein 27 - eq
seq_name: pir2:JG0168
seq_documentation_block:
gob-5 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
C:Accession: JG0168
R:Komiyama, T.; Tanigawa, Y.; Hirohashi, S.
Biochem. Biophys. Res. Commun. 255, 347-351, 1999
A:Title: Cloning and identification of the gene gob-5, which is expressed in intestin
A:Reference number: JG0168; MUID:99160866
A:Accession: JG0168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-913 <KOW>
A:Cross-references: DDBJ:AB016592
alignment_scores:
Quality: 313.00 Length: 72
Ratio: 4.408 Gaps: 0
Percent Similarity: 98.611 Percent Identity: 77.778
alignment_block:
US-09-049-696-9 x JG0168
Align seg 1/1 to: JG0168 from: 1 to: 913
2 AACATGGCGCTCATGTCATGCTTTGGGGCCCTTCATCAGGAATGGAGC 51
466 AsnAsnGlyLeuValAspAlaPheAlaLeuSerSerGlyAsnAlaL 482
52 TGTCTCTCAGCGCTCCATCAGCTTTCAGAGTAAAGGATTAACCTCCAGA 101
482 alleAlaGlnHisSerIleGlnLeuGluSerArgGlyValAsnLeuGlnA 499
102 ACAGCGATGGGATGAATGGCACAGTGCATGTCGGACAGCACCGTGGGAAG 151
499 snAsnGlnTrpMetAsnGlySerValIleValAspSerSerValGlyLys 515
152 GACACATTTGTTTTCATCAGCTCCGACAGCAGCCTCCCAATTCCTTCT 201
516 AspThrLeuPheLeuIleThrTrpThrHisProProThrIlePheI 532
202 CTGGGATCCAGTGGGA 217
532 eTrpAspProSerGly 537
seq_name: pir2:T02205
seq_documentation_block:
Lu-ECAM-1 protein - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 01-Dec-2000
C:Accession: T02205; T02152; T02171
R:Elble, R.C.; Widom, J.; Gruber, A.D.; Abdel-Ghany, M.; Levine, R.; Goodwin, A.; Pau
submitted to the EMBL Data Library, April 1997
A:Description: Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial
A:Reference number: Z14590
A:Accession: T02205
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-905 <ELB>
A:Cross-references: EMBL:AF001261; NID:g2623762; PIDN:AAB86529.1; PID:g2623763
A:Experimental source: lung
A:Accession: T02152
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA

164 GAAACAAGTGTCTCCTTTCCCAACGGTGTCTCCACGATCAGTGTGCCATTC 115

us-09-049-696-9.rpr

Tue Apr 2 09:40:16 2002

A:Residues: 1-450 <STO>
A:Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA06286.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: murD
C:Superfamily: UDP-N-acetylmuramate--alanine ligase

alignment_scores:
Quality: 65.50 Length: 72
Ratio: 1.637 Gaps: 3
Percent Similarity: 55.556 Percent Identity: 29.167

alignment_block:
US-09-049-696-9/rev x G83970 ..

Align seg 1/1 to: G83970 from: 1 to: 450

195 GATTGGGGAGGCTGGTTCAGGTGATAGAAA..... 161

30 AspleuGlyAlaIleValThrValAsnAspGlnLysProLeuAlaAspAs 46

160CAAGTGTCTTCCACGGTGTCTCCACGATCACTGTG 121

46 nProGlnAlaGlnLysLeuGlnLysGluGly...IleHisValValCysG 62

120 CCATTTCATCCAGTGGCTGTCTGGAGGGTTAATCCCTTACTCTCAAGCTG 71

62 lylGluHisProIleSerLeuLeuAspGlyLysGluLeuValLysAsn 78

70 GATGAGCGCTGAGACAGCTCATTTCTCTGATGAAGGCCGCCAAAAG 21

79 ProGlyIleArgTyrAspAsnProIle..... 87

20 CATCAATGAGGCCATT 5

88 ValGluGluAlaIle 92

seq_name: pirl:A42489

seq_documentation_block:
glutamate dehydrogenase (NADP+) (EC 1.4.1.4) - Giardia lamblia

N:Alternate names: glutamic dehydrogenase; NADP-specific glutamate dehydrogenase

C:Species: Giardia lamblia

C:Date: 31-Dec-1993 #sequence_revision 23-Mar-1995 #text_change 11-Jun-1999

C:Accession: A42489

J.Yee, J.; Dennis, P.P.
J. Biol. Chem. 267, 7539-7544, 1992

A:Title: Isolation and characterization of a NADP-dependent glutamate dehydrogenase

A:Reference number: A42489; MUID:92218410

A:Accession: A42489

A:Molecule type: DNA

A:Residues: 1-449 <YEE>

A:Cross-references: GB:M84604; NID:g159108; PIDN:AAA29155.1; PID:g159109

A>Note: sequence extracted from NCBI backbone (NCBIN:94071, NCBIP:94074)

C:Superfamily: glutamate dehydrogenase (NAD(P)+)

C:Keywords: NADP; oxidoreductase

F.125/Binding site: substrate (Lys) #status predicted

alignment_scores:
Quality: 64.50 Length: 47
Ratio: 2.016 Gaps: 2
Percent Similarity: 68.085 Percent Identity: 40.426

alignment_block:
US-09-049-696-9 x A42489 ..

Align seg 1/1 to: A42489 from: 1 to: 449

2 AACATGCGCTCATTTGGGCGCCTTCATCAGAAATGGAGC 51

223 AsnAsnThrValIleArgGlyLysAsnValLeuLeuSerGlySerGlyAs 239

.....
67 AspSerSerCysCysGlnProSerCysGluThrSerCysGlnPr 83
114 ATCC.....ACTGGCTGTCT...GGAGGGT 92
83 oSerCysTyrGlnThrSerSerCysGlyThrGlyCysGlyIleGlyGlyG 100

91 TAATCCCTTACTCTCAAGCTGGATGAGC 63

100 lyllegTyrGlnGlnGluGlySerSer 109

seq_name: pirl:2:38820

seq_documentation_block:
probable calcium-transporting atpase - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T38820

R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, September 1997

A:Reference number: Z21813

A:Accession: T38820

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1367 <CON>

A:Cross-references: EMBL:Z98980; NID:e1060691; PIDN:CA811719.1; GSPDB:GN000666; SPDB:SPAC

A:Experimental source: strain 972h-; cosmid c4F10

C:Genetics:

A:Gene: SPDB:SPAC4F10.16c

A:Map position: 1

A:Introns: 1176/2

alignment_scores:
Quality: 66.50 Length: 59
Ratio: 1.847 Gaps: 2
Percent Similarity: 61.017 Percent Identity: 32.203

alignment_block:
US-09-049-696-9 x T38820 ..

Align seg 1/1 to: T38820 from: 1 to: 1367

5 AATGGCTCATTCATCTTTGGGCGCCTTCATCAGGAAATGGAGCTGT 54

408 AsnGlyValLysAlaProGlyAlaValGlnSerProSerGluSerTh 424

55 CTCTCAGCGCTCCATCCAGCTTGAGAGT..... 82

424 RasnGlyArgLysIleHisGluGluProPheSerIleSerAsnValLeu 441

83 ..AAGGATTAAACCTCCAGACAGCCAGTGGATGAATGGCAGATGATC 130

441 euCysGlyCysThrLeuArgAsnSerLysTrpValIleGlyLeuValLeu 457

131 GTGACACACACCGCGGGAAGGACACT 157

458 Tyr.....ThrGlySerGluThr 463

seq_name: pirl:G83970

seq_documentation_block:
UDP-N-acetylmuramoylalanine-D-glutamate ligase murD [imported] - Bacillus halodurans (st

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000

C:Accession: G83970

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20263314

A:Accession: G83970

A:Status: preliminary

A:Molecule type: DNA

52 TGCTCTCAG.....CGTCCATCCAGCTTGAGCAAGGAT 89
 |||:||||| :|: |||:||||| :|:|||||
 239 nValAlaGlnPheAlaCysGluLysLeuIleGlnLeuGlyAlaLysVal 256
 |||:||||| :|: |||:||||| :|:|||||
 90 TAACCTCTCAGACAGCCAGCTGATGATGCGACAGTGATC 130
 |||:||||| :|: |||:||||| :|:|||||
 256 euThrPheSerAspSer.....AsnGlyThrIleVal 266

seq_name: pir2:I50106

seq_documentation_block:

MHC class I histocompatibility antigen alpha chain UAA*01 precursor - zebra fish (fragment)

C.Species: Brachydanio rerio (zebra fish)

C.Date: 13-Mar-1997 #sequence_revision 13-Mar-1997 #text_change 03-Dec-1999

C.Accession: I50106; S49816

R:Takeuchi, H.; Figueroa, F.; O'huigin, C.; Klein, J.

Immunogenetics 42, 77-84, 1995

A.Title: Cloning and characterization of class I Mhc genes of the zebrafish, Brachydanio

A.Reference number: I50106; MUID:95331830

A.Accession: I50106

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: mRNA

A.Residues: 1-229 <TAK>

A.Cross-references: EMBL:Z46776; NID:g577533; PIDN:CAA86731.1; PID:g577534

C.Genetics:

A.Gene: UAA*01

C.Superfamily: class I histocompatibility antigen; immunoglobulin homology

alignment_scores:
 Quality: 64.00 Length: 84
 Ratio: 1.488 Gaps: 5
 Percent Similarity: 51.190 Percent Identity: 26.190

alignment_block:

US-09-049-696-9 x I50106

Align seg 1/1 to: I50106 from: 1 to: 229

38 TCAGAAATGAGAGCTGCTCTCAGCGCTCCATCAGCTT..... 76
 |||:||||| :|: |||:||||| :|:|||||
 82 ThrGlyAlaHisProValPheLysAsnAsnIleGlnLysGluAr 98
 |||:||||| :|: |||:||||| :|:|||||

77GAGAGTAAGGATTAACCTCCAGACAGCCAGTGGATGATG 119

.....:||||| :|: |||:||||| :|:|||||

98 gPheAsnGlnThrGlnGly.....ValHisThrPheGlnTrpMetTyrG 113

|||:||||| :|: |||:||||| :|:|||||

120 GCACAGTATCGTGGACAGCACCGTG..... 145

|||:||||| :|: |||:||||| :|:|||||

113 lYcysGluLeuTyrAspAspGlyThrLysArgGlyTyrMetGlnTyrGly 129

|||:||||| :|: |||:||||| :|:|||||

146GGAAAGACACTTTGTTCTT.....ATCACCTCGAC 177

|||:||||| :|: |||:||||| :|:|||||

130 TyrAspGlyGluAspPheLeuSerLysLysThrLeuThrTrpTh 146

|||:||||| :|: |||:||||| :|:|||||

178 AACGCGACCTCCCAATCCTTCTC.....TGGGATCCCGAGTG 215

|||:||||| :|: |||:||||| :|:|||||

146 rAlaSerAsnProGlnAlaValIleThrLysValLysTrpAspSerThrG 163

|||:||||| :|: |||:||||| :|:|||||

216 GA 217

|||:||||| :|: |||:||||| :|:|||||

163 lY 163

seq_name: pir2:E83027

seq_documentation_block:

hypothetical protein PA4959 [imported] - Pseudomonas aeruginosa (strain PA01)

C.Species: Pseudomonas aeruginosa

C.Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C.Accession: E83027

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

.....:||||| :|: |||:||||| :|:|||||

213 CTGGATCCAGAGAGGATTTGGGAGGCTGCTGTTCAGGTGATAAG 164

|||:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A.Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
 A.Reference number: AB2950; MUID:20437337
 A.Accession: E83027
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-691 <STO>
 A.Cross-references: GB:AE004908; GB:AE004091; NID:g9951230; PIDN:AAG08344.1; GSPDB:CN
 A.Experimental source: strain PA01
 C.Genetics:
 A.Gene: PA4959

alignment_scores:
 Quality: 64.00 Length: 73
 Ratio: 1.455 Gaps: 6
 Percent Similarity: 60.274 Percent Identity: 35.616

alignment_block:

US-09-049-696-9 x E83027

Align seg 1/1 to: E83027 from: 1 to: 691

15 TTGATGCTTTTGGGCCCTTTCATCAGAAATGAGAGCTGTCTCTCAGCGC 64
 |||:||||| :|: |||:||||| :|:|||||

479 LeuArgLeuLeuAsnProGlnGlyGlnGluValProProAla.GluPheL 494

|||:||||| :|: |||:||||| :|:|||||

65 TCATCCAGCTTCGAGTAAGGATTAAACCTCCAGAACACGACCGATGGAT 114

|||:||||| :|: |||:||||| :|:|||||

494 euHisAlaAlaLysGluAlaGlyLeu.....AlaGluLysIleAsp 507

|||:||||| :|: |||:||||| :|:|||||

115 GAATGG.....CACATGATCGT.....GGACAGACCGTGGGAA 149

|||:||||| :|: |||:||||| :|:|||||

508 ArgTrpValIleLeuAsnSerIleLysLeuLeuAlaGluHisArgAlaLy 524

|||:||||| :|: |||:||||| :|:|||||

150 AGGACAC.....TTGTCTTCTTATCCTCGACACACGACCGCT 187

|||:||||| :|: |||:||||| :|:|||||

524 sGlyHisGlnThrLysLeuPheVal.....HisLeuSerSerAlaSerL 539

|||:||||| :|: |||:||||| :|:|||||

168 CCCAAATCCTCTCTG 204

|||:||||| :|: |||:||||| :|:|||||

539 euGlnAspProGlyLeu 544

|||:||||| :|: |||:||||| :|:|||||

seq_name: pir2:C40046

seq_documentation_block:

antibiotic transport-associated protein actII-3 - Streptomyces coelicolor

C.Species: Streptomyces coelicolor

C.Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 22-Oct-1999

C.Accession: C40046

R:Fernandez-Moreno, M.A.; Caballero, J.L.; Hopwood, D.A.; Malpartida, F.

Cell 66, 769-780, 1991

A.Title: The act cluster contains regulatory and antibiotic export genes, direct targ

A.Reference number: A40046; MUID:91347376

A.Molecule type: DNA

A.Residues: 1-711 <FER>

A.Cross-references: GB:M64683; NID:g153143; PIDN:AAA26691.1; PID:g153146

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

.....:||||| :|: |||:||||| :|:|||||

Tue Apr 2 09:40:16 2002

us-09-049-696-9.rpr

```

301 LeuAsnSerThrLysGlyLeuGlyProValCysAlaValGlyValLeuVa 317
163 AAACAAGTGCCTTCCACGGTGTCCACGATCAGTGTGCCATTCA 114
   ::::::::::::::::::::|||
317 lGlyLeuSerMetMetThrLeuLeuProAlaLeuValIlePheG 334
   ::::::::::::::::::::|||
113 TCCACTGGCTGTTCTGGAGGGTTAATCCC 85
   ::::::::::|||
334 lYargtrpValPheTrpProAlaArgPro 343
seq_name: pir2:S46795

seq_documentation_block:
hypothetical protein YHR002w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 23-Mar-2001
C:Accession: S46795
R:Favella, T.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of S. cerevisiae cosmid 9780.
A:Reference number: S46795
A:Accession: S46795
A:Molecule type: DNA
A:Residues: 1-357 <FAV>
A:Cross-references: EMBL:U10555; NID:g500813; PIDN:AB68424.1; PID:g500815; MIPS:YHR002w
C:Genetics:
A:Map position: 8R
A:Superfamily: ADP.ATP carrier protein; ADP.ATP carrier protein repeat homology
C:Keywords: duplication; mitochondrion; transmembrane protein
F:30-123/Domain: ADP.ATP carrier protein repeat homology <ACP1>
F:129-232/Domain: ADP.ATP carrier protein repeat homology <ACP2>
F:261-355/Domain: ADP.ATP carrier protein repeat homology <ACP3>

```

```

alignment_scores:
Quality: 63.50 Length: 70
Ratio: 1.549 Gaps: 3
Percent Similarity: 58.571 Percent Identity: 25.714

alignment_block:
US-09-049-696-9/rev x S46795 ..
Align seg 1/1 to: S46795 from: 1 to: 357

201 AGAAGATTGGGAGGCTGGTGTCCAGGTGATAAGAAACAAGTGTC 152
   ::::::::::::::::::::|||
167 LysArgValLysLeuGlyArgIleLysLysIleTyLysGluProAl 183
   ::::::::::::::::::::|||
151 CTTTCCACGGTGTCCACGATCAGTGTGCCA.....TTCATCCACT 108
   ::::::::::::::::::::|||
183 aSerAlaThrLeuIleLysAsnAspTyRileProAsnTrpPheCysHisT 200
   ::::::::::::::::::::|||
107 GGCTGTCTCGGA.....GGTTAATCCC 85
   ::::::::::|||
200 rp.CysAsnPhenylArgGlyTyRValProThrValLeuGlyMetIlePr 216
   ::::::::::|||
84 TTACTCTCAAGCTGGAGCGCTGAGACAGCTCCATTCTCTGATGA 35
   ::::::::::::::::::::|||
216 oTyRAlaGlyValSerPheAlaHisAspLeuLeuHisaspValLeuL 233
   ::::::::::::::::::::|||
34 AGGGCCC 27
   ::::::::::|||
233 ysSerPro 235

```

Database length: 22503292

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/pdata2/1/aa/5B COMB	pep:US-08-713-298B-2	+	64.50	127.61	2.95
/cgn2_6/pdata2/1/aa/5B COMB	pep:US-08-870-180B-2	+	64.50	127.61	2.95
/cgn2_6/pdata2/1/aa/6A COMB	pep:US-08-814-052-A	+	64.50	127.61	2.95
/cgn2_6/pdata2/1/aa/6A COMB	pep:US-08-812-829-A	+	64.50	127.61	2.95
/cgn2_6/pdata2/1/aa/6B COMB	pep:US-09-226-529-2	+	64.50	127.61	2.95
/cgn2_6/pdata2/1/aa/5B COMB	pep:US-08-870-180B-13	+	64.50	126.17	3.07
/cgn2_6/pdata2/1/aa/6B COMB	pep:US-09-226-529-13	+	64.50	126.17	3.07
/cgn2_6/pdata2/1/aa/5A COMB	pep:US-07-270-076A-11	+	61.50	113.60	8.82
/cgn2_6/pdata2/1/aa/5A COMB	pep:US-07-668-648-2	-	60.00	113.23	12.66
/cgn2_6/pdata2/1/aa/5B COMB	pep:US-08-429-998-2	-	60.00	113.23	12.66
/cgn2_6/pdata2/1/aa/5B COMB	pep:US-08-431-333-2	-	60.00	113.23	12.66
/cgn2_6/pdata2/1/aa/6A COMB	pep:PCT-US91-02321-2	-	60.00	113.23	12.66
/cgn2_6/pdata2/1/aa/6B COMB	pep:US-09-147-915-3	+	59.00	115.82	14.87
/cgn2_6/pdata2/1/aa/6B COMB	pep:US-08-639-294-2	+	59.00	115.82	14.87
/cgn2_6/pdata2/1/aa/6B COMB	pep:US-09-019-385-2	+	59.00	110.29	17.37
/cgn2_6/pdata2/1/aa/5A COMB	pep:US-08-653-740-3	+	58.00	108.75	22.92
/cgn2_6/pdata2/1/aa/5B COMB	pep:US-09-073-594-3	+	58.00	108.75	22.92
/cgn2_6/pdata2/1/aa/6A COMB	pep:US-09-275-925-3	+	58.00	108.75	22.92
/cgn2_6/pdata2/1/aa/5A COMB	pep:US-08-653-740-5	+	58.00	107.80	23.54
/cgn2_6/pdata2/1/aa/5B COMB	pep:US-09-073-594-5	+	58.00	107.80	23.54
/cgn2_6/pdata2/1/aa/6A COMB	pep:US-09-275-925-5	+	58.00	107.80	23.54
/cgn2_6/pdata2/1/aa/6B COMB	pep:US-08-653-740-3	+	58.00	107.80	23.54
/cgn2_6/pdata2/1/aa/5B COMB	pep:US-08-727-548-2	+	57.50	109.71	25.07
/cgn2_6/pdata2/1/aa/6B COMB	pep:US-09-340-637-61	-	57.00	108.73	30.65
/cgn2_6/pdata2/1/aa/5B COMB	pep:US-08-095-728B-6	-	57.00	108.73	30.65
/cgn2_6/pdata2/1/aa/6A COMB	pep:PCT-US92-02320A-6	-	57.00	106.73	30.65
/cgn2_6/pdata2/1/aa/5B COMB	pep:US-08-095-728B-2	-	57.00	103.21	33.84
/cgn2_6/pdata2/1/aa/6A COMB	pep:PCT-US92-02320A-2	-	57.00	103.21	33.84
/cgn2_6/pdata2/1/aa/5B COMB	pep:US-08-612-788-28	-	56.50	117.59	25.38
/cgn2_6/pdata2/1/aa/6A COMB	pep:US-09-066-028-28	-	56.50	117.59	25.38
/cgn2_6/pdata2/1/aa/5B COMB	pep:US-08-612-788-33	-	56.50	113.62	28.38
/cgn2_6/pdata2/1/aa/6A COMB	pep:US-09-066-028-33	-	56.50	113.62	28.38
/cgn2_6/pdata2/1/aa/5A COMB	pep:US-08-248-629A-6	-	56.50	110.58	30.91
/cgn2_6/pdata2/1/aa/5A COMB	pep:US-08-451-932-6	-	56.50	110.58	30.91
/cgn2_6/pdata2/1/aa/5A COMB	pep:US-08-452-260-6	-	56.50	110.58	30.91
/cgn2_6/pdata2/1/aa/5A COMB	pep:US-08-326-785-6	-	56.50	110.58	30.91
/cgn2_6/pdata2/1/aa/5B COMB	pep:US-08-612-788-6	-	56.50	110.58	30.91
/cgn2_6/pdata2/1/aa/5B COMB	pep:US-08-605-598B-6	-	56.50	110.58	30.91
/cgn2_6/pdata2/1/aa/5B COMB	pep:US-08-429-998-6	-	56.50	110.58	30.91

us-09-049-696-9.ra1

Tue Apr 2 09:40:15 2002

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-814-052-4

seq_documentation_block:
; Sequence 4, Application US/08814052
; Patent No. 6015783
; GENERAL INFORMATION:
; APPLICANT: von der Osten, Claus
; APPLICANT: Cherry, Joel R.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Vind, Jesper
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: PROCESS FOR REMOVAL OR BLEACHING OF SOILING
; TITLE OF INVENTION: OR STAINS FROM CELLULOSE FABRIC
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6015783o No. 6015783disk of No. 6015783th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/814.052
; FILING DATE: 06-MAR-1997
; CLASSIFICATION: 510
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4684.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-814-052-4

alignment_scores:
Quality: 64.50 Length: 47
Ratio: 2.016 Gaps: 1
Percent Similarity: 68.085 Percent Identity: 31.915

align seg 1/1 to: US-08-814-052-4 from: 1 to: 400
2 AACATGGCCCTCATTCATGCTTTTGGGGCCCTTCATCAGGAATGGA... 49
|||||
23 AsnAspSerValValGluGluHisGlyGlnLeuSerIleSerAsnGlyG1 44
50 .GCTGTCTCTCAGCGCTCCATCCAGCTTCAGAGTAGAGGATTAACCCCTCC 98
|||||
44 uLeuValAsnGluArgGlyGluGlnValGlnLeuLysGlyMetSerSerH 61
99 AGAACACCCAGTGGATGAATGGCACAGTGCCTGGACAGC 139
|||||
61 isGlyLeuGlnTrpTyrGlyGlnPheValAsnTyrGluSer 74

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-812-829-4
seq_documentation_block:

61 isGlyLeuGlnTrpTyrGlyGlnPheValAsnTyrGluSer 74

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-870-180B-2

seq_documentation_block:
; Sequence 2, Application US/08870180B
; Patent No. 5945327
; GENERAL INFORMATION:
; APPLICANT: Outtrup, Helle
; APPLICANT: Dammann, Claus
; APPLICANT: Olsen, Arne
; APPLICANT: Bisg rd-Frantzen, Henrik
; APPLICANT: Sch lein, Martin
; APPLICANT: J rgensen, Per
; APPLICANT: Bjoernvad, Mads
; TITLE OF INVENTION: DNA Constructs and Methods of Producing
; TITLE OF INVENTION: Cellulytic Enzymes
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5945327o No. 5945327disk of No. 5945327th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentio Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,180B
; FILING DATE: 6-June-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol E.
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 3794.434-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-870-180B-2

alignment_scores:
Quality: 64.50 Length: 47
Ratio: 2.016 Gaps: 1
Percent Similarity: 68.085 Percent Identity: 31.915

align seg 1/1 to: US-08-870-180B-2 from: 1 to: 400
2 AACATGGCCCTCATTCATGCTTTTGGGGCCCTTCATCAGGAATGGA... 49
|||||
28 AsnAspSerValValGluGluHisGlyGlnLeuSerIleSerAsnGlyG1 44
50 .GCTGTCTCTCAGCGCTCCATCCAGCTTCAGAGTAGAGGATTAACCCCTCC 98
|||||
44 uLeuValAsnGluArgGlyGluGlnValGlnLeuLysGlyMetSerSerH 61
99 AGAACACCCAGTGGATGAATGGCACAGTGCCTGGACAGC 139
|||||
61 isGlyLeuGlnTrpTyrGlyGlnPheValAsnTyrGluSer 74

CORRESPONDENCE ADDRESS:
ADDRESSSEE: NO. 59453270 No. 5945327disk of No. 5945327th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/870.180B
FILING DATE: 6-June-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol E.
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 3794.434-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-870-180B-13

```

alignment_scores:      Length: 47
                        Quality: 64.50
                        Ratio: 2.016
                        Gaps: 1
                        Percent Identity: 31.915

alignment_block:
  US-09-049-696-9 x US-09-226-529-13 ..
  Align seg 1/1 to: US-09-226-529-13 from: 1 to: 462

      2 AACAAATGGCCTCATGTGCTTTTGGGCCCTTTCATCAGAAATGGA.. 49
      |||.....:||||| ||| ||| ||| ||| ||| ||| ||| |||
      23 AsnAspSerValValGluGluHisGlyGlnLeuSerIleSerAsnGlyG1 44
      |||.....:||||| ||| ||| ||| ||| ||| ||| ||| |||
      50 .GCNCTCTCTCAGCGCCCATCCACGCTTGAGAGTAAGGGATTAACCCCTCC 98
      |||.....:||||| ||| ||| ||| ||| ||| ||| ||| |||
      44 uLeuValAsnGluargGlyGluGlnValGlnLeuLysGlyMetSerH 61
      |||.....:||||| ||| ||| ||| ||| ||| ||| ||| |||
      99 AGAACAGCCAGTGGATGAATGGCAGCATGATCGTGGACAGC 139
      |||.....:||||| ||| ||| ||| ||| ||| ||| ||| |||
      61 IsGlyLeuGlnTrpTyrGlyGlnPheValAsnTyrGluSer 74

seq_name: /cgn2_6/prodata2/iaa/5A_COMB.pep:US-08-270-076A-11

seq_documentation_block:
  : Sequence 11, Application US/08270076A
  : Patent No. 5667986
  : GENERAL INFORMATION:
  : APPLICANT: Sleep, Darrell
  : APPLICANT: Goodey, Andrew R
  : APPLICANT: Vakeria, Diana
  : TITLE OF INVENTION: Yeast Promoter

```


;; TITLE OF INVENTION: MODULATING PROTEINS
;; NUMBER OF SEQUENCES: 12
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Pennie & Edmonds
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10036
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/429,998
;; FILING DATE: 27-APR-1995
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/668,648
;; FILING DATE: 13-MAR-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Misrock, S. Leslie
;; REGISTRATION NUMBER: 18,872
;; REFERENCE/DOCKET NUMBER: 5624-161-999
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 790-9090
;; TELEFAX: (212) 869-9741
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 589 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-429-998-2

alignment_scores:
Quality: 60.00 Length: 55
Ratio: 2.143 Gaps: 4
Percent Similarity: 50.909 Percent Identity: 29.091

alignment_block:
US-09-049-696-9/rev x US-08-429-998-2 ..

Align seg 1/1 to: US-08-429-998-2 from: 1 to: 589

```
155 TGCTCTTTCCACGG.....TGCTGTCC 133
|||||  ::::|
305 CysProPheThrLysAlaValCysCysGluAspHisIleHisCysCysPr 321
132 ACGATCACTG...TGCCATTTCACCTGGC...TGTTCTGGAGGGTTAA 89
| ::::| |||||  ::::|
321 oAlaGlyPheGlnCysHisThrGlyThrCysGluLeuGlyVal 338
88 TCCTTACTCTCAAGCTGAGCGCTGAGACAGCTCCATTTCCTG 39
:::  |||  :::
338 eu.....GlnValProTrrpMetLysLysValThrAlaSerLeuSer 351
38 ATGAAAGGGCCCA 24
:::  |||||
352 LeuProAspProGln 356
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pap:US-08-431-333-2

seq_documentation_block:
; Sequence 2, Application US/08431333
; Patent No. 5965723
; GENERAL INFORMATION:
; APPLICANT: Shoyab, Mohammed
; APPLICANT: Plowman, Gregory D.
; TITLE OF INVENTION: NOVEL CYSTEINE-RICH GROWTH
; MODULATING PROTEINS

;; NUMBER OF SEQUENCES: 12
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Pennie & Edmonds
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10036
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/431,333
;; FILING DATE: 27-APR-1995
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/668,648
;; FILING DATE: 13-MAR-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Misrock, S. Leslie
;; REGISTRATION NUMBER: 18,872
;; REFERENCE/DOCKET NUMBER: 5624-161-999
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 790-9090
;; TELEFAX: (212) 869-9741
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 589 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-431-333-2

alignment_scores:
Quality: 60.00 Length: 55
Ratio: 2.143 Gaps: 4
Percent Similarity: 50.909 Percent Identity: 29.091

alignment_block:
US-09-049-696-9/rev x US-08-431-333-2 ..

Align seg 1/1 to: US-08-431-333-2 from: 1 to: 589

```
155 TGCTCTTTCCACGG.....TGCTGTCC 133
|||||  ::::|
305 CysProPheThrLysAlaValCysCysGluAspHisIleHisCysCysPr 321
132 ACGATCACTG...TGCCATTTCACCTGGC...TGTTCTGGAGGGTTAA 89
| ::::| |||||  ::::|
321 oAlaGlyPheGlnCysHisThrGlyThrCysGluLeuGlyVal 338
88 TCCTTACTCTCAAGCTGAGCGCTGAGACAGCTCCATTTCCTG 39
:::  |||  :::
338 eu.....GlnValProTrrpMetLysLysValThrAlaSerLeuSer 351
38 ATGAAAGGGCCCA 24
:::  |||||
352 LeuProAspProGln 356
```

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pap:PCT-US91-02321-2

seq_documentation_block:
; Sequence 2, Application PC/TUS9102321
; GENERAL INFORMATION:
; APPLICANT: Shoyab, Mohammed
; APPLICANT: Plowman, Gregory D.
; TITLE OF INVENTION: NOVEL CYSTEINE-RICH GROWTH
; MODULATING PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:

```
; ADDRESSEE: Bristol-Myers Squibb Company
; STREET: 3005 First Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT-US91/02321
; FILING DATE: 19910403
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Poor, Brian W.
; REGISTRATION NUMBER: 32,928
; REFERENCE/DOCKET NUMBER: ONO071A-PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)728-4800
; TELEFAX: (206)448-4775
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 589 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US91-02321-2

alignment_scores:
    Quality:        60.00      Length:      55
                   Ratio:     2.143          Gaps:       4
Percent Similarity: 50.909   Percent Identity: 29.091

alignment_block:
US-09-049-696-9/rev x PCT-US91-02321-2 ..

Align seg 1/1 to: PCT-US91-02321-2 from: 1 to: 589

155 TGTCCTTTCCACGG.....TGCTGTCC 133
||||||| :||| |||||
305 CysProPhrThrLysAlaValCysGluAspHisIleHiscysCysPr 321
132 ACGATCACTG...TGCATTATCATCCACTGCC...TCCTCTGGAGGTTAA 89
+ : : : : + : : : : : : : : : | : : : : | : : : :
321 oAglYpheIncYsiHsthrGluthrGlyThrCysGluLeuGlyVall 338

88 TCCCTTACTCTCAAGCTGGATGGAGCGCTGAGACAGACGTCCATCCCTCG 39
||||| :||| :|||
338 eu.....GlnValProTrpMetLysLysValThrAlaSerLeuSer 351

38 ATGAAGGGCCCCNA 24
::: |||||
352 LeuProAspProGln 356

seq_name: /cqn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-147-915-3

seq_documentation_block:
; Sequence 3, Application US/09147915A
; Patent No. 6184034
; GENERAL INFORMATION:
; APPLICANT: Eastman, Alan
; TITLE OF INVENTION: Deoxyribonuclease II Proteins and cDNAs
; FILE REFERENCE: DC-0097
; CURRENT APPLICATION NUMBER: US/09/147,915A
; EARLIER FILING DATE: 1999-03-23
; CURRENT APPLICATION NUMBER: PCT/US97/18262
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/028,539
; EARLIER FILING DATE: 1996-10-15
```

us-09-049-696-9-rai

Tue Apr 2 09:40:15 2002

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3562
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 360 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-639-294-2

alignment_scores:
Quality: 59.00 Length: 87
Ratio: 1.475 Gaps: 3
Percent Similarity: 45.977 Percent Identity: 22.989

alignment_block:
US-09-049-696-9 x US-08-639-294-2 ..
Align seg 1/1 to: US-08-639-294-2 from: 1 to: 360
41 GGAATGGAGCTCTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATT 90
||||| : : : : : : : : : : : : : : : : : : : : : : : :
40 GlySerGlyGluAlaAlaGlnArgGlyLeuGlnTyr.....LysTyrLe 54
| : : : : : : : : : : : : : : : : : : : : : : : :
91 AACCTCCAGAACAGCCAGTGGATGATGAATGCACAGTGTGTCGACAGCA 140
| : : : : : : : : : : : : : : : : : : : : : : : :
54 uAspGluSerSerGlyGlyTrpArgAspGlyArgAlaLeuIleAsnSerP 71
| : : : : : : : : : : : : : : : : : : : : : : : :
141 CCCTGGGA.....AAG 151
|||
71 roGluGlyAlaValGlyArgSerLeuGlnProLeuTyrArgSerAsnThr 87
152 GACACTTGTGTTTATCCTACCTGGACACAGCAGCTCCCAA..... 193
: : : : : : : : : : : : : : : : : : : : : : : :
88 SerGlnLeuAlaPheLeuLeuTyrAsnAspGlnProGlnProSerLy 104
| : : : : : : : : : : : : : : : : : : : : : : : :
194 .....ATCCTTCTCTGGG 206
| : : : : : : : : : : : : : : : : : : : : : : : :
104 sAlaGlnAspSerSerMetArgGlyHisThrLysGlyValLeuLeuA 121
207 ATCCAGTGA 217
|| : : : :
121 sPHisAspGly 124

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-019-385-2

seq_documentation_block:
; Sequence 2, Application US/09019385
; Patent No. 6147280
; GENERAL INFORMATION:
; APPLICANT: Smeekens, Josephus C.M.
; APPLICANT: Ebskamp, Michael J.M.
; APPLICANT: Geerts, Hendrikus A.M.
; APPLICANT: Weisbeek, Petrus J.
; TITLE OF INVENTION: Production of oligosaccharides in
; TITLE OF INVENTION: Transgenic Plants
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Christensen O'Connor Johnson & KindnessPLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,385
```

seq_name: /SIDS2/qcgcdata/geneseq/geneseqp/AA2001.DAT: AAG73854

seq_documentation_block:

US	homo sapiens:	
XX	1	
XX	2	
PN	WO200122920-A2.	
XX		
XX		
PD	05-APR-2001.	
XX		
XX		
PF	28 ⁵ -SEP-2000; 2000WO-US26524.	
XX		
XX		
PR	29-SEP-1999; 99US-0157137.	
FR	03-NOV-1999; 99US-0163280.	
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAG73894 +	/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAG73894 +	374.00	839.34	1.0e-39	552
/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAG75614 +	/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAG75614 +	374.00	843.34	1.0e-39	869
/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB74824 +	/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB74824 +	374.00	838.86	1.0e-39	914
/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB748514 +	/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB748514 +	374.00	838.86	1.0e-39	914
/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB73716 +	/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB73716 +	374.00	838.86	1.0e-39	914
/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB74733 +	/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB74733 +	374.00	838.86	1.0e-39	914
/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAG75474 +	/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAG75474 +	374.00	838.74	1.0e-39	925
/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB74822 +	/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB74822 +	313.00	697.12	8.1e-32	913
/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB73715 +	/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB73715 +	313.00	697.12	8.1e-32	913
/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB766749 +	/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB766749 +	291.00	645.94	5.7e-29	919
/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB87560 +	/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB87560 +	291.00	645.94	5.7e-29	919
/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB865272 +	/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB865272 +	291.00	645.94	5.7e-29	919
/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAG53310 +	/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAG53310 +	280.00	634.09	1.1e-27	218
/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAG737671 +	/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAG737671 +	280.00	634.09	1.1e-27	218
/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB741039 +	/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB741039 +	187.50	409.64	1.3e-15	592
/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB71324 +	/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB71324 +	187.50	409.64	1.3e-15	592
/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB845904 +	/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB845904 +	187.50	407.48	1.4e-15	742
/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB741040 +	/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB741040 +	187.50	406.87	1.4e-15	791
/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB11325 +	/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB11325 +	187.50	406.87	1.4e-15	791
/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB770459 +	/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB770459 +	187.50	405.21	1.4e-15	942
/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB74823 +	/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB74823 +	187.50	405.20	1.4e-15	943
/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB710361 +	/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB710361 +	187.50	405.20	1.4e-15	943
/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB11321 +	/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB11321 +	187.50	405.20	1.4e-15	943
/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB73171 +	/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB73171 +	69.50	134.34	2.49	665
/SIDS2/cgdata/geneseq/geneseq/AA1987.DAT.AAB70183 +	/SIDS2/cgdata/geneseq/geneseq/AA1987.DAT.AAB70183 +	67.50	137.50	3.76	293
/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AA735222 +	/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AA735222 +	67.00	127.86	5.32	713
/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB73848 +	/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB73848 +	66.50	141.75	4.34	147
/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB869365 +	/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB869365 +	65.50	135.71	6.38	217
/SIDS2/cgdata/geneseq/geneseq/AA1997.DAT.AAM23601 +	/SIDS2/cgdata/geneseq/geneseq/AA1997.DAT.AAM23601 +	64.50	127.56	9.85	400
/SIDS2/cgdata/geneseq/geneseq/AA1997.DAT.AAM22521 +	/SIDS2/cgdata/geneseq/geneseq/AA1997.DAT.AAM22521 +	64.50	127.56	9.85	400
/SIDS2/cgdata/geneseq/geneseq/AA1998.DAT.AAM57431 +	/SIDS2/cgdata/geneseq/geneseq/AA1998.DAT.AAM57431 +	64.50	127.56	9.85	400
/SIDS2/cgdata/geneseq/geneseq/AA1998.DAT.AAM57433 +	/SIDS2/cgdata/geneseq/geneseq/AA1998.DAT.AAM57433 +	64.50	126.19	10.17	462
/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB541920 +	/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB541920 +	64.00	139.24	8.47	104
/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB11338 +	/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AAB11338 +	63.50	153.79	6.82	20
/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB202455 +	/SIDS2/cgdata/geneseq/geneseq/AA2001.DAT.AAB202455 +	63.50	155.66	16.59	1093
/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB74828 +	/SIDS2/cgdata/geneseq/geneseq/AA1999.DAT.AAB74828 +	63.00	150.89	8.24	24
/SIDS2/cgdata/geneseq/geneseq/AA2000.DAT.AA744384 +	/SIDS2/cgdata				

```
alignment_scores:
  Quality: 374.00
  Ratio: 5.194
  Percent Similarity: 100.000
  Length: 72
  Gaps: 0
  Percent Identity: 100.000

alignment_block:
  US-09-049-696-g x RAG73854
```

CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention. 682 and page 7053 of the sequence listing were
CC N.B. Pages 686 to 692 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ. ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 869 AA;

Align seg 1/1 to: AAG73854 from: 1 to: 552

2 AACATGGCTCATGATGCTTTGGGGCCCTTTCATCAGGAATGGAGC 51
|||||
103 AsnAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAl 119
52 TGCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAACCCCTCCAGA 101
|||||
119 aValSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnA 136
102 ACAGCCAGTGGATGAATGCGACAGTATCGTGGACAGCACCGTGGGAAG 151
|||||
136 snSerGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLys 152
152 GACACTTTGTTCTTATACCTCGGACAGCGACGCTCCGCCAAATCCTTCT 201
|||||
153 AspThrLeuPheLeuIleThrThrThrGlnProProGlnIleLeuLe 169
202 CTGGGATCCAGTGA 217
|||||
169 uTrpAspProSerGly 174

seq_name: /SDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAG75614

seq_documentation_block:
ID AAG75614 standard; Protein; 869 AA.

XX
AC AAG75614;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6378.
XX
KW Human: colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WC200122920-A2.

XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
XX
PS N-PSDB; AAH35019.

XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 7851-7854; 9803pp; English.

XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis

alignment_scores:
Quality: 374.00 Length: 72
Ratio: 5.194 Caps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-9 x AAG75614

Align seg 1/1 to: AAG75614 from: 1 to: 869

2 AACATGGCTCATGATGCTTTGGGGCCCTTTCATCAGGAATGGAGC 51
|||||
420 AsnAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAl 436
52 TGCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAACCCCTCCAGA 101
|||||
436 aValSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnA 453
102 ACAGCCAGTGGATGAATGCGACAGTATCGTGGACAGCACCGTGGGAAG 151
|||||
453 snSerGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLys 469
152 GACACTTTGTTCTTATACCTCGGACAGCGACGCTCCGCCAAATCCTTCT 201
|||||
470 AspThrLeuPheLeuIleThrThrThrGlnProProGlnIleLeuLe 486
202 CTGGGATCCAGTGA 217
|||||
486 uTrpAspProSerGly 491

seq_name: /SDS2/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74824

seq_documentation_block:
ID AAB74824 standard; Protein; 914 AA.

XX
AC AAB74824;
XX
DT 13-JUN-2001 (first entry)
XX
DE Human ICACC-1 protein sequence.

XX
KW ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW inflammatory bowel disease; autoimmune disease.

XX
OS Homo sapiens.
XX
PN WC9944620-A1.
XX
PD 10-SEP-1999.

XX
PF 03-MAR-1999; 99WO-US04703.
XX
PR 03-MAR-1998; 98US-0076815.
XX
PS (MAGA-) MAGALIN PHARM INC.

XX
PI Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;
PI Nicolaides NC, Zhou Y, Dong Q;
XX
DR WPI; 1999-550979/46.

DR N-PSDB; AAF81927.
 XX New nucleic acid encoding calcium activated chloride channel, used to
 PT identify, e.g. specific modulators for treating atopic allergy -
 XX
 XX
 PS Claim 11; Fig 4B; 75pp; English.
 XX
 XX The present sequence represents the human interleukin 9 (IL-9) induced
 CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
 CC have anti-allergic, anti-asthmatic, anti-inflammatory and
 CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
 CC used to alleviate asthma (or more generally atopic allergy), while those
 CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
 CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
 CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
 CC (increased levels) or IBD (reduced levels), also for monitoring
 CC treatment of these conditions. The ICACC proteins can be used:
 CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
 CC reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
 CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
 CC identify modulators and binding partners. ICACC polynucleotides can be
 CC used to generate transgenic animals or recombinant cells, used to screen
 CC for antagonists, also as a source of therapeutic antisense agents or
 CC diagnostic probes (for quantifying mRNA expression, e.g. for
 CC identification of modulators).
 XX
 SQ Sequence 914 AA;

alignment_scores:
 Quality: 374.00 Length: 72
 Ratio: 5.194 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-049-696-9 x AAB74824

Align seg 1/1 to: AAB74824 from: 1 to: 914

2 AACAAATGGCCCTCATGATGCTTTTGGGGCCCTTTTCATCAGGAATAGGAGC 51
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 465 ASnAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAl 481
 52 TGCTCTCTCAGCGCTCATCCAGCTGAGCTAAGGGATTAAACCTCCAGA 101
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 481 aValSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnA 498
 102 ACAGCAGTGGATGAATGGCACAGTGTGTCGACAGCACCGTGGGAAAG 151
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 498 snSerGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLys 514
 152 GACACATTTGTTCTTATCAGCTGGACACAGCGACCGCTCCCGCAATCCTTCT 201
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 515 AspThrLeuPheLeuIleThrTrpThrThrGlnProGlnIleLeuLe 531

202 CTGGGATCCAGTGA 217
 ||||||||||||||||
 531 utrpAspprosergly 536

seq_name: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT:AA24514

seq_documentation_block:

ID AM24514 standard; Protein; 914 AA.

XX

AC AM24514;

XX

DT 12-OCT-2001 (first entry)

XX

DE C902P predicted amino acid sequence.

XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;

KW immunogenic; gene therapy; vaccine; colonic cancer.

XX

OS Homo sapiens.
 XX
 PN WO200149716-A2.
 XX
 PD 12-JUL-2001.
 XX
 XX 29-DEC-2000; 2000WO-US35596.
 PF
 XX 30-DEC-1999; 99US-0476296.
 PR 10-JAN-2000; 2000US-0480321.
 PR 15-FEB-2000; 2000US-0504629.
 PR 06-MAR-2000; 2000US-0519444.
 PR 19-MAY-2000; 2000US-0575251.
 PR 29-JUN-2000; 2000US-0609448.
 PR 28-AUG-2000; 2000US-0649811.
 XX
 XX (GRI-) CORIXA CORP.
 XX
 XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
 Z1 King GE, Wang T, Jiang Y;
 XX
 DR WPI; 2001-441847/47.
 XX
 XX Colon tumor associated proteins and nucleic acids useful for the
 PT prevention, diagnosis and treatment of colonic cancer -
 XX
 PS
 XX
 XX Claim 2; Page 440-443; 472pp; English.
 XX
 XX The present invention describes colon tumour associated proteins (I) and
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumour associated protein (TCAP)
 CC expression, such as colonic cancer. For example, (I) and (II) may be
 CC used to treat disorders associated with decreased expression by
 CC rectifying mutations or deletions in a patient's genome that affect the
 CC activity of TCAPs by expressing inactive proteins or to supplement the
 CC patients own production of them. Additionally, (II) may be used to
 CC produce the TCAP proteins, by inserting the nucleic acids into a host
 CC cell culturing the cell to express the protein. (II) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and
 CC quantitate the presence of similar nucleic acids in samples, and
 CC therefore which patients may be in need of restorative therapy. (I) may
 CC also be used as antigens in the production of antibodies against TCAPs
 CC and in assays to identify modulators of TCAP expression and activity.
 CC Anti-(I) antibodies and antagonists may also be used to down regulate
 CC TCAP expression and activity. The anti-(I) antibodies may also be used
 CC as diagnostic agents for detecting the presence of TCAPs in samples
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AA128460 to AA129512
 CC and AA24494 to AA24523 represent nucleotide and amino acid sequences
 XX given in the exemplification of the present invention.
 XX
 SQ Sequence 914 AA;

alignment_scores:
 Quality: 374.00 Length: 72
 Ratio: 5.194 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-9 x AAM24514

Align seg 1/1 to: AAM24514 from: 1 to: 914

2 AACAAATGGCCCTCATGATGCTTTTGGGGCCCTTTTCATCAGGAATAGGAGC 51
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 465 ASnAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAl 481
 52 TGCTCTCTCAGCGCTCATCCAGCTGAGCTAAGGGATTAAACCTCCAGA 101
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 481 aValSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnA 498

```
102 ACAGCCAGTGGATGATGGACACAGTGTGCGGACAGCAGCGTGGGAAAG 151
|||||
498 snserGlntrpMetAsnGlyThrValIleValAspSerThrValGlyLys 514
|||||
152 GACACTTTGTTTTCATCCTGACGACAGCGAGCGTCCCAATCCTTCT 201
|||||
515 AspThrLeuPheLeuIleThrTrpThrThrGlnProGlnIleLeuLe 531

202 CTGGGATCCAGTGA 217
|||||
531 utrAspProSerGly 536

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB73716
seq_documentation_block:
ID AAB73716 standard; Protein: 914 AA.
XX AC AAB74733;
XX DT 12-JUN-2001 (first entry)
XX DE Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.
XX KW Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
XX KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;
XX KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
XX KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
XX KW antialzheimer; antiparkinsonian; antimicrobial; vulnary; gene therapy;
XX KW immune disorder; hyperproliferative disorder; cardiovascular disease;
XX KW cancer; angiogenic disorder; neurological disorder; infectious disease;
XX KW wound healing; regeneration; chemotaxis; chromosome 1.
XX CS Homo sapiens.
XX FN WO200112775-A2.
XX PD 22-FEB-2001.
XX PF 16-AUG-2000; 2000WO-US22325.
XX PR 17-AUG-1999; 99US-0149182.
XX PI (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ni J, Florence KA, Fiscella M, Wei P, Baker KP;
XX PI Birse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;
XX DR WFL: 2001-147550/15.
XX DR N-PSDB; AAF81787.
XX PT Nucleic acids encoding 25 human secreted polypeptides, useful for
XX PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX PT disease and diabetic retinopathy -
XX PS Claim 11; Page 459-460; 485pp; English.
XX CC AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733
XX CC to AAB74772. Human secreted proteins can have activities based on the
XX CC tissues and cells they are expressed in. Example of activities include:
```

```
alignment_block:
US_09-049-696-9 x AAB73716
Align seg 1/1 to: AAB73716 from: 1 to: 914
2 AACAAATGGCCATTTGATGCTTTGGGCGCCCTTTTCATCAGGAAATGGAGC 51
|||||
465 AsnAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAl 481
|||||
52 TGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAACCCCTCCAGA 101
|||||
481 aValSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnA 498
|||||
102 ACAGCCAGTGGATGATGGACACAGTGTGCGGACAGCAGCGTGGGAAAG 151
|||||
498 snserGlntrpMetAsnGlyThrValIleValAspSerThrValGlyLys 514
|||||
152 GACACTTTGTTTTCATCCTGACGACAGCGAGCGTCCCAATCCTTCT 201
|||||
515 AspThrLeuPheLeuIleThrTrpThrThrGlnProGlnIleLeuLe 531
|||||
202 CTGGGATCCAGTGA 217
|||||
531 utrAspProSerGly 536

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB74733
seq_documentation_block:
ID AAB74733 standard; Protein: 914 AA.
XX AC AAB74733;
XX DT 12-JUN-2001 (first entry)
XX DE Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.
XX KW Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
XX KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;
XX KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
XX KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
XX KW antialzheimer; antiparkinsonian; antimicrobial; vulnary; gene therapy;
XX KW immune disorder; hyperproliferative disorder; cardiovascular disease;
XX KW cancer; angiogenic disorder; neurological disorder; infectious disease;
XX KW wound healing; regeneration; chemotaxis; chromosome 1.
XX CS Homo sapiens.
XX FN WO200112775-A2.
XX PD 22-FEB-2001.
XX PF 16-AUG-2000; 2000WO-US22325.
XX PR 17-AUG-1999; 99US-0149182.
XX PI (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ni J, Florence KA, Fiscella M, Wei P, Baker KP;
XX PI Birse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;
XX DR WFL: 2001-147550/15.
XX DR N-PSDB; AAF81787.
XX PT Nucleic acids encoding 25 human secreted polypeptides, useful for
XX PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX PT disease and diabetic retinopathy -
XX PS Claim 11; Page 459-460; 485pp; English.
XX CC AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733
XX CC to AAB74772. Human secreted proteins can have activities based on the
XX CC tissues and cells they are expressed in. Example of activities include:
```

```
102 ACAGCCAGTGGATGATGGACACAGTGTGCGGACAGCAGCGTGGGAAAG 151
|||||
498 snserGlntrpMetAsnGlyThrValIleValAspSerThrValGlyLys 514
|||||
152 GACACTTTGTTTTCATCCTGACGACAGCGAGCGTCCCAATCCTTCT 201
|||||
515 AspThrLeuPheLeuIleThrTrpThrThrGlnProGlnIleLeuLe 531

202 CTGGGATCCAGTGA 217
|||||
531 utrAspProSerGly 536

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB73716
seq_documentation_block:
ID AAB73716 standard; Protein: 914 AA.
XX AC AAB73716;
XX DT 11-SEP-2001 (first entry)
XX DE Human CLCAL protein, SEQ ID NO:2.
XX KW Human CLCAL; goblet cell; mouse Gob-5 orthologue; drug screening;
XX KW expression inhibition; antisense therapy; gene therapy;
XX KW chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Misc-difference 152
XX FT /note= "Encoded by AGG in AAH46124"
XX PN WO200138530-A1.
XX PD 31-MAY-2001.
XX PF 22-NOV-2000; 2000WO-JP08232.
XX PR 24-NOV-1999; 99JP-0333479.
XX PR 27-APR-2000; 2000JP-0127589.
XX PA (TAKE ) TAKEDA CHEM IND LTD.
XX PI Nakanishi A, Morita S;
XX PI WPI: 2001-355935/37.
XX DR N-PSDB; AAH46102, AAH46124.
XX PT New antisense nucleotide, useful for treatment and prevention of
XX PT bronchial asthma and chronic obstructive pulmonary disease -
XX PS Claim 2; Page 76-80; 104pp; Japanese.
XX CC The invention relates to an antisense nucleotide targetted to the mouse
XX CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
XX CC the CLCAL gene (coding sequence shown in AAH46102). The invention also
XX CC relates to an antibody specific for the Gob-5 protein, medical and
XX CC diagnostic compositions containing the antisense nucleotide or the
XX CC antibody, and methods and kits for screening for compounds which inhibit
XX CC the protein. Gob-5 and CLCAL are proteins expressed by goblet cells.
XX CC The antisense oligonucleotides and antibody are therefore useful for the
XX CC treatment and prevention of bronchial asthma and chronic obstructive
XX CC pulmonary disease. The present sequence represents human CLCAL protein.
XX SQ Sequence 914 AA;
```

alignment_scores:
Quality: 374.00 Length: 72
Ratio: 5.194 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;
CC antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;
CC vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic;
CC anticonvulsant; antialzheimers; antiparkinsonian; antimicrobial; and
CC vulnery. Human secreted proteins can be used in gene therapy and
CC vaccine. Human secreted protein nucleotide sequences (NAMI) and proteins
CC (PEP1) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. For example, NAMI
CC and PEP1 may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patients genome
CC that affect the activity of proteins by expressing inactive proteins or
CC to supplement the patients own production of polypeptides. Disorders that
CC may be prevented, diagnosed and/or treated include immune disorders,
CC hyperproliferative disorders (e.g. cancers), cardiovascular diseases,
CC angiogenic disorders, neurological disorders, infectious diseases and/or
CC for promoting wound healing, regeneration and/or chemotaxis. AAF81778 to
CC AAF81786 and AAB74732 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 914 AA;

alignment_scores:
Quality: 374.00 Length: 72
Ratio: 5.194 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-9 x AAB74733

Align seg 1/1 to: AAB74733 from: 1 to: 914

2 ACAATGGCCTCATTTGATGCTTTTGGGCGCCTTTCATCAGGAATGGAGC 51
|||||
465 AsnAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAl 481
52 TGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGGATTAAACCTCCAGA 101
|||||
481 aValSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnA 498
102 ACAGCCAGTGGATGAATGGCACAGTGTGAGAGTGGACAGCAGCGTGGGAAG 151
|||||
498 snSerGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLys 514
152 GACACTTTGTTCTTATCAGCTGGACACAGCAGCCTCCCAATCTTCT 201
|||||
515 AspThrLeuPheLeuIleThrTrpThrThrGlnProGlnIleLeuLe 531
202 CTGGATCCAGTGA 217
|||||
531 uTrpAspProSerGly 536

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAG75474

seq_documentation_block:

ID AAG75474 standard; Protein; 925 AA.

XX
AC AAG75474;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6238.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 1.
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
DR N1PSDB; AAB34879.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
PT claim 11; Page 7686-7690; 9803pp; English.
XX
CC AAB32943 to AAB37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAB37196 to AAB37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 925 AA;

alignment_scores:
Quality: 374.00 Length: 72
Ratio: 5.194 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-9 x AAG75474

Align seg 1/1 to: AAG75474 from: 1 to: 925

12 ACAATGGCCTCATTTGATGCTTTTGGGCGCCTTTCATCAGGAATGGAGC 51
|||||
476 AsnAsnGlyLeuIleAspAlaPheGlyAlaLeuSerSerGlyAsnGlyAl 492
52 TGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGGATTAAACCTCCAGA 101
|||||
492 aValSerGlnArgSerIleGlnLeuGluSerLysGlyLeuThrLeuGlnA 509
102 ACAGCCAGTGGATGAATGGCACAGTGTGAGAGTGGACAGCAGCGTGGGAAG 151
|||||
509 snSerGlnTrpMetAsnGlyThrValIleValAspSerThrValGlyLys 525
152 GACACTTTGTTCTTATCAGCTGGACACAGCAGCCTCCCAATCTTCT 201
|||||
526 AspThrLeuPheLeuIleThrTrpThrThrGlnProGlnIleLeuLe 542
202 CTGGATCCAGTGA 217
|||||
542 uTrpAspProSerGly 547

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74822

seq_documentation_block:

ID AAB74822 standard; Protein; 913 AA.

XX AC AAB74822;
 XX 13-JUN-2001 (first entry)
 XX Murine ICACC-1 protein sequence.
 DE ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
 XX interleukin 9 induced calcium activated chloride channel; IL-9;
 KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
 KW anti-inflammatory; immunomodulatory; cystic fibrosis;
 KW inflammatory bowel disease; autoimmune disease.
 XX Mus sp.
 OS
 XX WO9944620-A1.
 XX 10-SEP-1999.
 XX 03-MAR-1999; 99WO-US04703.
 XX 03-MAR-1998; 98US-0076815.
 XX (MAGA-) MAGAININ PHARM INC.
 XX Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;
 PI Nicolaidis NC, Zhou Y, Dong Q;
 XX WPI; 1999-550979/46.
 DR N-PSDB; AAF81925.
 XX New nucleic acid encoding calcium activated chloride channel, used to
 PT identify, e.g. specific modulators for treating atopic allergy -
 XX Claim 12; Fig 2; 75pp; English.
 XX The present sequence represents the murine interleukin 9 (IL-9) induced
 CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
 CC have anti-allergic, anti-asthmatic, anti-inflammatory and
 CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
 CC used to alleviate asthma (or more generally atopic allergy), while those
 CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
 CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
 CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
 CC (increased levels) or IBD (reduced levels), also for monitoring
 CC treatment of these conditions. The ICACC proteins can be used:
 CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
 CC reagents, and (b) as therapeutic (AT); (ii) as fragments or mutants, as
 CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
 CC identify modulators and binding partners. ICACC polynucleotides can be
 CC used to generate transgenic animals or recombinant cells, used to screen
 CC for antagonists, also as a source of therapeutic antisense agents or
 CC identification of modulators).
 XX Sequence 913 AA;
 alignment_scores:
 Quality: 313.00 Length: 72
 Ratio: 4.408 Gaps: 0
 Percent Similarity: 98.611 Percent Identity: 77.778
 alignment_block:
 US-09-049-696-9 x AAB74822 ..
 Align seg 1/1 to: AAB74822 from: 1 to: 913
 2 AACAATGGCCTCATTGATGCTTTTGGGCCCTTTTCATCAGGAATGGAGC 51
 |||||
 466 AsnAsnGlyLeuValAspAlaPheAlaLeuSerSerGlyAsnAlaAl 482
 52 TGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAAGGATTAAACCTCCAGA, 101

482 atleAlaGlnHisSerIleGlnLeuGluSerArgGlyValAsnLeuGlnA 499
 |||||
 102 ACAGCCAGTGGATGAATGGACACAGTATCGTGGACAGCACACCGTGGAAAG 151
 |||||
 499 snAsnGlnTrpMetAsnGlySerValIleValAspSerValGlyLys 515
 |||||
 152 GACACTTTGTTCTTATCATCTGGACAGCGACGCTCCCAATCCTTCT 201
 |||||
 516 AspThrLeuPheLeuIleThrTrpThrHisProProThrIlePheIle 532
 |||||
 202 CTCGGATCCCAAGTGA 217
 |||||
 532 enrAspProSerGly 537
 seq_name: /SID32/gcgdata/geneseq/geneseq/AA2001.DAT:AAB73715
 seq_documentation_block:
 ID AAB73715 standard; Protein; 913 AA.
 XX
 AC AAB73715;
 XX
 DT 11-SEP-2001 (first entry)
 XX
 DE Mouse Gob-5 protein, SEQ ID NO:1.
 XX
 KW Mouse Gob-5; murine; goblet cell; human CLCA1 orthologue; drug screening;
 KW expression inhibition; antisense therapy; gene therapy; bronchial asthma;
 KW chronic obstructive pulmonary disease; antiasthmatic.
 XX
 OS Mus sp.
 XX WO200138530-A1.
 XX 31-MAY-2001.
 XX 22-NOV-2000; 2000WO-JP08232.
 XX 24-NOV-1999; 99JP-0333479.
 XX 27-APR-2000; 2000JP-0127589.
 XX (TAKE) TAKEDA CHEM IND LTD.
 XX Nakanishi A, Morita S;
 XX WFI; 2001-355935/37.
 XX N-PSDB; AAH46101, AAH46120.
 XX New antisense nucleotide, useful for treatment and prevention of
 XX bronchial asthma and chronic obstructive pulmonary disease -
 XX Claim 1; Page 72-76; 104pp; Japanese.
 XX The invention relates to an antisense nucleotide targetted to the mouse
 CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
 CC the CLCA1 gene (coding sequence shown in AAH46102). The invention also
 CC relates to an antibody specific for the Gob-5 protein, medical and
 CC diagnostic compositions containing the antisense nucleotide or the
 CC antibody, and methods and kits for screening for compounds which inhibit
 CC the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.
 CC The antisense oligonucleotides and antibody are therefore useful for the
 CC treatment and prevention of bronchial asthma and chronic obstructive
 CC pulmonary disease. The present sequence represents mouse Gob-5 protein.
 XX
 XX Sequence 913 AA;
 alignment_scores:
 Quality: 313.00 Length: 72
 Ratio: 4.408 Gaps: 0
 Percent Similarity: 98.611 Percent Identity: 77.778
 alignment_block:
 US-09-049-696-9 x AAB74822 ..

US-09-049-696-9 x AAB73715 ..
Align seg 1/1 to: AAB73715 from: 1 to: 913
2 AACAAATGCGCTCATATGATGCTTTTGGGGCCCTTTTCATCAGGAATGAGC 51
|||||
466 AsnAsnGlyLeuValAspAlaPheAlaAlaLeuSerSerGlyAsnAlaAl 482
52 TGCTCTCAGCGCTCCATCCAGCTTGAGTAGAGGATTAAACCTCCAGA 101
|||||
482 aileAlaGlnHisSerIleGlnLeuGluSerArgGlyValAsnLeuGlnA 499
102 ACACGACGATGATCAATGGCACAGTGCATGCGGACACACCGTGGAAAG 151
|||||
499 snAsnGlnIrrMetAsnGlySerValIleValAspSerSerValGlyLys 515
152 GACACTTGTGTTCTTATCCTACCTGGACACAGCAGCTCCCAATCCTTCT 201
|||||
516 AspThrLeuPheLeuIleThrTrpThrHisProProThrIlePheII 532
202 CTGGGATCCAGTGA 217
:|||||
532 etipAspProSerGly 537

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAV66749

seq_documentation_block:

ID AAV66749 standard; protein; 919 AA.

XX AC AAV66749;

XX DT XX

XX DE 05-APR-2000 (first entry)

XX KW Membrane-bound protein PRO1124.

XX KW Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;

XX KW pharmaceutical; receptor immunoadhesin; gene mapping.

XX OS Homo sapiens.

XX PN WO9963088-A2.

XX PD 09-DEC-1999.

XX PF 02-JUN-1999; 99WO-US12252.

XX PR 02-JUN-1998; 98US-0087607.

XX PR 02-JUN-1998; 98US-0087609.

XX PR 03-JUN-1998; 98US-0087759.

XX PR 04-JUN-1998; 98US-0087827.

XX PR 04-JUN-1998; 98US-0088021.

XX PR 04-JUN-1998; 98US-0088025.

XX PR 04-JUN-1998; 98US-0088028.

XX PR 04-JUN-1998; 98US-0088030.

XX PR 04-JUN-1998; 98US-0088033.

XX PR 05-JUN-1998; 98US-0088326.

XX PR 05-JUN-1998; 98US-0088167.

XX PR 05-JUN-1998; 98US-0088202.

XX PR 05-JUN-1998; 98US-0088212.

XX PR 05-JUN-1998; 98US-0088217.

XX PR 09-JUN-1998; 98US-0088655.

XX PR 10-JUN-1998; 98US-0088722.

XX PR 10-JUN-1998; 98US-0088730.

XX PR 10-JUN-1998; 98US-0088734.

XX PR 10-JUN-1998; 98US-0088738.

XX PR 10-JUN-1998; 98US-0088740.

XX PR 10-JUN-1998; 98US-0088741.

XX PR 10-JUN-1998; 98US-0088742.

XX PR 10-JUN-1998; 98US-0088810.

XX PR 10-JUN-1998; 98US-0088811.

XX PR 10-JUN-1998; 98US-0088824.

XX PR 10-JUN-1998; 98US-0088825.

PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.

US-09-049-696-9 x AAY66749 ..
Align seg 1/1 to: AAY66749 from: 1 to: 919
12 AACATGGCTCATTCATGCTTTTGGGCCCCCTTCATCAGGAATGGAGC 51
466 AsnAsnGlyLeuIleAspAlaPheGlyAlaLeuThrSerGlyAsnThrAs 482
52 TGTCCTCTCAGCGCTCCATCCAGCTGAGAGTAAGGATTAACCCCTCCAGA 101
482 pleuSerGlnLysSerLeuGlnLeuGluSerLysGlyLeuThrLeuAsnS 499
102 ACAGCCAGTGGATGAATGCACAGTCATCGTGCAGCAGCACCGTGGGAAG 151
499 erAsnAlaTrpMetAsnAspThrValIleAspSerThrValGlyLys 515
152 GACACTTTGTTCTTATCACCTGCACAGCAGCCTCCCAATCCTCTCT 201
516 AspThrPheLeuIleThrTrpAsnSerLeuProProSerIleSerLe 532
202 CTGGGATCCAGTGA 217
532 utrAspProSerGly 537
seq_name: /SDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA87560
seq_documentation_block:
ID AA87560 standard; Protein; 919 AA.
XX
AC AA87560;
XX
DT 15-MAY-2001 (first entry)
XX
DE Human PRO1124.
XX
KW Human; PRO protein; mapping.
XX
OS Homo sapiens.
XX
PN WC200116318-A2.
XX
PD 08-MAR-2001.
XX
PF 24-AUG-2000; 2000WO-US23328.
XX
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 07-DEC-1999; 99US-0169495.
PR 09-DEC-1999; 99US-0170262.
PR 11-JAN-2000; 2000US-0175481.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 01-MAR-2000; 2000WO-US05601.
PR 03-MAR-2000; 2000US-0187202.
PR 25-APR-2000; 2000US-0199397.
PR 22-MAY-2000; 2000WO-US14042.
PR 05-JUN-2000; 2000US-0209832.
XX
XX (GETH) GENENTECH INC.
XX
XX Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
PI N-PSDB; AAF92092.
XX
DR WPI; 2001-183260/18.
DR N-PSDB; AAF92092.
XX
XX Eighty four nucleic acids encoding PRO polypeptides, useful in
PT molecular biology, including use as hybridization probes, and in
PT chromosome and gene mapping.
XX
PS Claim 12; Fig 70; 278pp; English.
XX

10-AUG-1998; 98US-0095929.
10-AUG-1998; 98US-0096012.
11-AUG-1998; 98US-0096143.
11-AUG-1998; 98US-0096146.
12-AUG-1998; 98US-0096329.
17-AUG-1998; 98US-0096757.
17-AUG-1998; 98US-0096768.
17-AUG-1998; 98US-0096773.
17-AUG-1998; 98US-0096791.
17-AUG-1998; 98US-0096867.
17-AUG-1998; 98US-0096891.
17-AUG-1998; 98US-0096894.
17-AUG-1998; 98US-0096895.
17-AUG-1998; 98US-0096897.
18-AUG-1998; 98US-0096949.
18-AUG-1998; 98US-0096950.
18-AUG-1998; 98US-0096959.
18-AUG-1998; 98US-0096960.
18-AUG-1998; 98US-0097022.
18-AUG-1998; 98US-0097141.
20-AUG-1998; 98US-0097218.
24-AUG-1998; 98US-0097661.
26-AUG-1998; 98US-0097951.
26-AUG-1998; 98US-0097952.
26-AUG-1998; 98US-0097954.
26-AUG-1998; 98US-0097955.
26-AUG-1998; 98US-0097971.
26-AUG-1998; 98US-0097974.
26-AUG-1998; 98US-0097978.
26-AUG-1998; 98US-0097979.
26-AUG-1998; 98US-0097986.
26-AUG-1998; 98US-0098014.
31-AUG-1998; 98US-0098525.
16-SEP-1998; 98US-0100634.
12-JAN-1999; 99US-0115565.
XX
XX (GETH) GENENTECH INC.
XX
XX Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
XX WPI; 2000-072883/06.
XX N-PSDB; AA265095.
XX
XX Membrane-bound proteins and related nucleotide sequences
XX
XX Claim 12; Fig 274; 822pp; English.
XX
XX The invention provides membrane-bound PRO polypeptides and
CC polynucleotides encoding them. The PRO sequences of the invention were
CC identified based on extracellular domain homology screening. The PRO
CC sequences have homology with proteins including LDL receptors, TIE
CC molecules and various enzymes. The membrane-bound proteins and receptor
CC ligands are useful as pharmaceutical and diagnostic agents. Receptor
CC immunoadhesins, for instance, can be used as therapeutic agents to block
CC receptor-ligand interactions. The membrane-bound proteins can also be
CC employed for screening of potential peptide or small molecule inhibitors
CC of the relevant receptor/ligand interaction. The PRO encoding sequences
CC are useful as hybridization probes, in chromosome and gene mapping and in
CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC will also be useful for the preparation of PRO polypeptides, especially
CC by recombinant techniques.
XX
XX Sequence 919 AA;
SQ

alignment_scores:
Quality: 291.00 Length: 72
Ratio: 4.409 Gaps: 0
Percent Similarity: 91.667 Percent Identity: 75.000

alignment_block:

152 GACACTTGTCTTATCACCTGGACACGAGCGCTCCCAATCCTTCT 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
516 AspThrPheLeuIleThrTrpAsnSerLeuProSerIleSerLe 532
202 CTGGGATCCAGTGA 217
532 utrPaspProSerGly 537
seq_name: /SIDS2/cgdata/geneseq/geneseq/AA2000.DAT:AA53310
seq_documentation_block:
ID AAB53310 standard; Protein; 218 AA.
AC AAB53310;
DT 09-MAR-2001 (first entry)
DE Human colon cancer antigen protein sequence SEQ ID NO:850.
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.
XX Homo sapiens.
XX WO200055351-A1.
XX 21-SEP-2000.
XX 08-MAR-2000; 2000WO-US05883.
XX 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM;
XX WPI; 2000-587534/55.
XX N-PSDB; AAC98067.
XX Colon cancer associated gene sequences, referred to as colon cancer
XX antigens, useful for the treatment, prevention, and diagnosis of colon
XX disorders such as colon cancer -
XX Claim 11; Page 1400; 2104pp; English.
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
XX called human colon cancer antigens, given in AAB53234 to AAB54006. The
XX human colon cancer antigens can have cytostatic, cardioactive, muscular;
XX neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
XX vulnerary, nephrotropic, antiinfective and antibacterial activities, and
XX can be used in gene therapy. The colon cancer antigen polynucleotides,
XX proteins and antibodies to the proteins are useful for the prevention,
XX treatment and diagnosis of colon disorders, such as colon cancer. The
XX polynucleotides may be used in diagnostics and research, such as for
XX chromosome identification, and as hybridisation probes. The proteins
XX may also be used to prevent diseases such as neural disorders, immune
XX system disorders, muscular disorders, reproductive disorders,
XX gastrointestinal disorders, wounds, renal disorders, infectious
XX diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
XX AAB54007 represent sequences used in the exemplification of the present
XX invention.
XX Sequence 218 AA;
SQ
alignment_scores:
Quality: 280.00 Length: 72
Ratio: 4.375 Gaps: 0
Percent Similarity: 88.889 Percent Identity: 72.222
alignment_block:
US-09-049-696-9 x AAB53310 ..
Align seg 1/1 to: AAB53310 from: 1 to: 218
2 ACAATGGCTCATTCATGCTTTTGGGCGCTTTCATCAGGAATGAGC 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
85 AsnAsnGlyLeuIleAspAlaPheGlyAla***ThrSerGlyAsnThrAs 102
52 TGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAACCTCCACA 101
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
102 pLeuSer***LysSerLeuGlnLeuGluSerLysGlyLeuThrLeuAsnS 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
102 ACACCCAGTGGATGAATGGCACAGATGATCGTGCACACACCGTGGGAAAG 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
119 erAsnAlaIrrpMetAsnAspThrValIleAspSerThrValGlyLys 135
152 GACACTTGTCTTATCACCTGGACACGAGCGCTCCCAATCCTTCT 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
136 AspThrPheLeuIleThrTrpAsnSerLeuProSerIleSerLe 152
202 CTGGGATCCAGTGA 217
152 utrPaspProSerGly 157
seq_name: /SIDS2/cgdata/geneseq/geneseq/AA2001.DAT:AA673761
seq_documentation_block:
ID AAG73761 standard; Protein; 218 AA.
AC AAG73761;
XX 03-SEP-2001 (first entry)
XX Human colon cancer antigen protein SEQ ID NO:4525.
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma.
XX Homo sapiens.
XX WO200122920-A2.
XX 05-APR-2001.
XX 28-SEP-2000; 2000WO-US26524.
XX 29-SEP-1999; 99US-0157137.
XX 03-NOV-1999; 99US-0163280.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
XX N-PSDB; AAB33192.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 11; Page 6321; 9803pp; English.
XX AAB32943 to AAB37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions

CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 218 AA;

alignment_scores:
Quality: 280.00 Length: 72
Ratio: 4.375 Gaps: 0
Percent Similarity: 88.889 Percent Identity: 72.222

alignment_block:
US-09-049-696-9 x AAG73761

Align seg 1/1 to: AAG73761 from: 1 to: 218

2 AACAAATGCGCTCATTTGATGCTTTTGGGGCCCTTTTCATCAGGAATGGAGC 51
|||||
86 AsnAsnGlyLeuIleAspAlaPheGlyAla***ThrSerGlyAsnThrAs 102
|||||
52 TGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAACCTCCAG 101
|||||
102 pLeuSer***LysSerLeuGlnLeuGluSerLysGlyLeuThrLeuAsnS 119
|||||
102 ACACCCAGTGGATGAATGGGACAGTATCGTGGACAGCACCGTGGGAAAG 151
|||||
119 erAsnAlaTrpMetAsnAspThrValIleIleAspSerThrValGlyLys 135
|||||
152 GACACTTTGTTCTTATCACCCTGACACAGCAGCTCCCAAAATCCTTCT 201
|||||
136 AspThrPhePheLeuIleThrTrpAsnSerLeuProProSerIleSerLe 152
|||||
202 CTGGGATCCCAAGTGA 217
|||||
152 uTrpAspProSerGly 157

seq_name: /SDS2/gcgdata/geneseq/geneseq/AA1999.DAT.AAY41039

seq_documentation_block:

ID AAY41039 standard; protein; 592 AA.

XX AAY41039;

XX 07-DEC-1999 (first entry)

XX Human lung tumor antigen L762P variant 1.

XX Human; lung tumor; lung cancer; T cell stimulation.

XX Homo sapiens.

XX WO9947674-A2.

XX 23-SEP-1999.

XX 17-MAR-1999; 99WO-US05798.

XX 18-MAR-1998; 98US-0040802.

XX 18-MAR-1998; 98US-0040984.

XX 27-JUL-1998; 98US-0123912.

XX 27-JUL-1998; 98US-0123933.

XX (CORI-) CORIXA CORP.

PI Reed SG, Wang T;
XX
XX WPI; 1999-571839/48.
XX N-PSDB; AA224657.
XX
XX New isolated lung tumor polynucleotides, used to develop products for
XX the treatment, prevention and monitoring the progression of lung cancer
XX
XX Example 3; Page 144-145; 148pp; English.
XX
XX The invention provides isolated human lung tumor nucleic acids and
XX polypeptides. The polypeptides can be used for the treatment of lung
XX cancer. The polypeptides and polynucleotides can be used to stimulate T
XX cells or antigen presenting cells for use in the treatment of lung
XX cancer. The polypeptides and monoclonal antibodies specific for the
XX polypeptides can also be used to inhibit the development of lung cancer.
XX Agents which bind the polypeptides can be used for detecting lung cancer
XX and for monitoring the progression of lung cancer.

SQ Sequence 592 AA;

alignment_scores:

Quality: 187.50 Length: 73
Ratio: 3.074 Gaps: 1
Percent Similarity: 83.562 Percent Identity: 49.315

alignment_block:

US-09-049-696-9 x AAY41039

Align seg 1/1 to: AAY41039 from: 1 to: 592

2 AACAAATGCGCTCATTTGATGCTTTTGGGGCCCTTTTCATCAGGAATGGAGC 51
|||||
473 SerAsnSerMetIleAspAlaPheSerArgIleSerSerGlyThrGlyAs 489
|||||
52 TGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAACCTCCAGA 101
|||||
489 pIlePheGlnGlnHisIleGlnLeuGluSerThrGlyGluAsnValLysp 506
|||||
102 ACAGCCAGTGGATGAATGGGACAGTATCGTGGACAGCACCGTGGGAAAG 151
|||||
506 roHisHisGlnLeuLysAsnThrValThrValAspAsnThrValGlyAsn 522
|||||
152 GACACTTTGTTCTTATCACCCTGG...ACAACGACGCTCCCAAAATCCT. 198
|||||
523 AspThrMetPheLeuValThrTrpGlnAlaSerGlyProProGluIleI 539
|||||
199 TCTCTGGATCCCAAGTGA 217
|||||
539 eLeuPheAspProAspGly 545

Tue Apr 2 09:40:15 2002

us-09-049-696-9.rag

This Page Blank (uspto)

OM of: US-09-049-696-8 to: SPTREMBL_17.* out_format : pfs

Date: Mar 30, 2002 2:46 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+nt2p.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09049696/runat_28032002_145238_2053/app_query.fasta_1.12579
-DB=SPTREMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=bloms62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696_eCGN1_1_705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-8
Query length: 253
Database: SPTREMBL_17.*
Database sequences: 473505
Database length: 146272329
Search time (sec): 805.760000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
sp_human:Q9UNF7	+	276.00	586.66	2.3e-25	917	Q9UNF7 homo sapiens (human). ca
sp_rodent:Q9R070	+	240.00	507.28	6.2e-21	902	Q9R070 mus musculus (mouse). ca
sp_rodent:Q88860	+	239.00	505.08	8.2e-21	901	Q88860 mus musculus (mouse). ch
sp_rodent:Q9QX15	+	235.00	505.07	8.2e-21	902	Q9QX15 mus musculus (mouse). ca
sp_rodent:Q9EQ84	+	235.00	496.23	2.5e-20	902	Q9EQ84 mus musculus (mouse). en
sp_human:Q9NXP1	+	222.00	472.63	1.0e-18	469	Q9NXP1 homo sapiens (human). cd
sp_mammal:O18742	+	216.00	455.27	5.5e-18	794	Q18742 bos taurus (bovine). lu
sp_mammal:O18743	+	216.00	455.02	5.5e-18	820	Q18743 bos taurus (bovine). lu
sp_mammal:O18741	+	216.00	454.25	5.5e-18	905	Q18741 bos taurus (bovine). lu
sp_human:Q9UQC9	+	194.00	405.34	2.8e-15	943	Q9UQC9 homo sapiens (human). ch
sp_human:Q9Y6N2	+	194.00	405.34	2.8e-15	943	Q9Y6N2 homo sapiens (human). ca
sp_invertebrate:Q77441	+	85.00	175.62	0.0718	230	Q77441 giardia lamblia (giard
sp_invertebrate:Q97449	+	85.00	175.62	0.0718	230	Q97449 giardia lamblia (giard
sp_invertebrate:Q9TVV3	+	85.00	175.62	0.0718	230	Q9TVV3 giardia lamblia (giard
sp_invertebrate:Q9XTJ8	+	83.00	171.20	0.1266	230	Q9XTJ8 giardia lamblia (giard
sp_invertebrate:Q9U045	+	82.00	170.21	0.1680	197	Q9U045 giardia lamblia (giard
sp_invertebrate:Q27871	+	82.00	169.00	0.1681	230	Q27871 giardia lamblia (giard
sp_invertebrate:Q9U047	+	82.00	169.00	0.1681	230	Q9U047 giardia lamblia (giard
sp_invertebrate:Q9U046	+	77.00	159.12	0.6925	198	Q9U046 giardia lamblia (giard
sp_invertebrate:Q27663	+	77.00	157.95	0.6927	230	Q27663 giardia lamblia (giard
sp_bacteria:Q69822	+	76.50	138.81	0.8017	2314	Q69822 streptomyces coelicolor
sp_bacteria:Q9K084	+	74.00	135.19	1.63	1815	Q9K084 neisseria meningitidis
sp_bacteria:Q44911	+	73.50	148.28	1.87	295	Q44911 borrelia burgdorferi (ly
sp_bacteria:Q9K3V1	+	72.50	147.33	2.48	251	Q9K3V1 streptomyces coelicolor.
sp_mammal:Q9GM48	+	71.00	140.88	3.79	375	Q9GM48 macaca fascicularis (cra
sp_human:Q9BXX1	+	69.50	135.45	5.81	492	Q9BXX1 homo sapiens (human). an
sp_bacteria:Q55113	+	69.00	139.63	6.68	250	Q55113 synechocystis sp. (strai
sp_bacteria:Q06454	+	69.00	136.57	6.69	370	Q06454 neisseria meningitidis.
sp_bacteria:Q06467	+	68.00	136.57	6.69	370	Q06467 neisseria gonorrhoeae.
sp_bacteria:Q9AIU6	+	68.00	123.16	8.90	1552	Q9AIU6 neisseria meningitidis.
sp_bacteria:Q9PEY9	+	68.00	117.31	8.91	3282	Q9PEY9 xylella fastidiosa. hem
sp_bacteria:Q9PBE8	+	68.00	116.94	8.91	3442	Q9PBE8 xylella fastidiosa. hem
sp_bacteria:Q9P9U6	+	68.00	116.91	8.91	3455	Q9P9U6 xylella fastidiosa. hem
sp_bacteria:Q53348	+	67.50	125.07	10.25	1055	Q53348 mycobacterium tubercul
sp_invertebrate:Q9W117	+	67.50	124.77	10.25	1096	Q9W117 drosophila melanogast
sp_plant:Q9FXJ8	+	67.50	121.32	10.25	1705	Q9FXJ8 arabidopsis thaliana (m
sp_bacteria:Q84900	+	67.00	135.67	11.77	236	Q84900 escherichia coli. struct
sp_bacteria:Q06468	+	67.00	133.19	11.78	334	Q06468 neisseria meningitidis.
sp_fungi:Q36028	+	67.00	121.94	11.81	1367	Q36028 schizosaccharomyces pom

seq_name: sp_human.Q9UNF7

seq_documentation_block:

ID	Q9UNF7	PRELIMINARY	PRT	917 AA.
AC	Q9UNF7			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 2.			
GN	CAC2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
CX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=COLON;			
RX	MEDLINE=99364503; PubMed=10437792;			
RA	Aguel M., Vermet T., Culouscou J.M.;			
RT	"Identification of three novel members of the calcium-dependent			
RT	chloride channel (cacc) family predominantly expressed in the			
RT	digestive tract and trachea."			
RL	FEBS Lett. 455:295-301(1999).			
DR	EMBL: AF127035; AAD48398.1; .			
DR	InterPro: IPR002035; VWFA.			
DR	Pfam: PF00092; vwa; 1.			
DR	PROSITE: PS50234; VWFA; 1.			
DR	SMART: SM00327; VWA; 1.			
SQ	SEQUENCE 917 AA; 101153 MW; EA01090D781BEB95 CRC64;			

alignment_scores:

Quality	276.00	Length	83
Ratio	3.833	Gaps	0
Percent Similarity	86.747	Percent Identity	63.855
US-09-049-696-8 x Q9UNF7			
Align seg 1/1 to: Q9UNF7 from: 1 to: 917			
3 CAAAGTGGTGCATCCACACACAGCTGGGCGCCTCTGCAGCTCA 52			
423 GlnSerGlyAlaIleValHisPheIleAlaLeuGlyArgAlaAspGI 444			
53 AGAAGTACAGAGAGCTGCCAAATGACAGAGGTTTACAGACATATGCTT 102			
444 uAlaValIleGluMetSerIleThrGlySerHisPheYrValS 461			
103 CAGATCAAGTTCAGAACAAATGGCCCTCATTTGATGCTTTGGGGCCCTTCA 152			
461 erAspGluAlaGlnAsnAsnGlyLeuIleAspAlaPheGlyAlaLeuThr 477			
153 TCAGAAATGAGCTGCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGG 202			
473 SerGlyAsnThrAspLeuSerGlnIlySerLeuGlnLeuIlySerLysGI 494			
203 ATTAACCTCCAGAACACGCCAGTGGATGATGGACAGTGTGCTGGAC 251			
494 yLeuThrLeuAsnSerAsnAlaTrpMetAsnAspThrValIleIleAsp 510			

seq_name: sp_rodent:Q9R070

seq_documentation_block:

ID	Q9R070	PRELIMINARY	PRT	902 AA.
AC	Q9R070			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	CA(2+)-SENSITIVE CHLORIDE CHANNEL 2.			
GN	CLCA2 OR CACC.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

us-09-049-696-8.ispt

Tue Apr 2 09:40:15 2002

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20012773; PubMed=10544033;
RT Lee D., Ha S., Kho Y., Kim J., Cho K., Baik M., Choi Y.;
RT "Induction of mouse Ca(2+)-sensitive chloride channel 2 gene during
RT involution of mammary gland.";
RL Biochem. Biophys. Res. Commun. 264:933-937(1999).
DR EMBL: AF108501; AAF12731.1; -.
DR MGD: MGI:1931471; Clca2.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 901 AA; 100038 MW; 13DF068BF3337AAB CRC64;

alignment_scores:
Quality: 239.00 Length: 83
Ratio: 3.366 Gaps: 1
Percent Similarity: 85.542 Percent Identity: 59.036

alignment_block:
US-09-049-696-8 x 088860
Align seg 1/1 to: 088860 from: 1 to: 901
3 CAAGTGGTGCATCATCCACAGCTCGCTTTGGGGCCCTCTGCAGCTCA 52
430 ArgSerGlyAlaIleIleHisThrIleAlaLeuGlyProSerAlaAa 446
53 AGAAGTACAGAGAGCTGTCACAAATCAGAGAGGTTTACAGACATATGCTT 102
446 gGluLeuGluThrLeuSerAspMetThrGlyLeuArgPheTyAlaA 463
103 CAGATCAAGTTCAGAACAAATGCCTCATGATGCTTTGGGGCCCTTCA 152
463 snlyAspLeu.....AsnSerLeuIleAspAlaPheSerArgIleSer 477
153 TCAGGAATGGAGCTGCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGG 202
478 SerThrSerGlySerValSerGlnGlnAlaLeuGlnLeuGluSerLysAl 494
203 ATTAACCTCCAGAACAGCCAGTGGATGAATGCACAGTATGCTGGGAC 251
494 apheAsnValArgAlaGlyAlaTripleAsnGlyThrValProLeuAsp 510

seq_name: sp_rodent:Q9QX15
seq_documentation_block:
ID Q9QX15 PRELIMINARY; PRT; 902 AA.
AC Q9QX15;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE CALCIUM-SENSITIVE CHLORIDE CONDUCTANCE PROTEIN-1.
GN CLCA1 OR MCLCA1
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-WHOLE LUNG;
RX MEDLINE=99041980; PubMed=9822685;
RA Gandhi R., Eble R.C., Gruber A.D., Schreier K.D., Ji H.-L.,
RA Fuller C.M., Pauli B.U.;
RT "Molecular and functional characterization of a calcium-sensitive
RT chloride channel from mouse lung";
RL J. Biol. Chem. 273:32096-32101(1998).
DR EMBL: AF047838; AAC79982.1; -.
DR MGD: MGI:1316732; Clca1.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 902 AA; 100107 MW; D9A915E698EA2D85 CRC64;

alignment_scores:
Quality: 239.00 Length: 83
Ratio: 3.366 Gaps: 1
Percent Similarity: 85.542 Percent Identity: 59.036

alignment_block:
US-09-049-696-8 x'Q9QX15
Align seg 1/1 to: Q9QX15 from: 1 to: 902

```

```

3 CAAAGTGGTCCCATCATCCACAGTCGGCTTTGGGGCCCTCTCGAGCTCA 52
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
430 ArgSerGlyAlaIleIleHisThrIleAlaLeuGlyProSerAlaAlaAr 446
53 AGAAGTAGAGGAGCTGCCAAATGACAGAGGTTTACACACATATGCTT 102
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
446 gLuLeuGluThrLeuSerAspMetThrGlyGlyLeuArgPheThrAlaA 463
103 CAGATCAAGTTCAGAACAAATGGCTCATTCATGCTTTGGGGCCCTTCA 152
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
463 snlyAspLeu.....AsnSerLeuIleAspAlaPheSerArgIleSer 477
153 TCAGGAATGGAGCTGTCTTCAGCGCTCCATCCAGCTTGAGAGTAAGGG 202
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
478 SerThrSerGlySerValSerGlnGlnAlaLeuGlnLeuGluSerIysAl 494
203 ATTAACCCCTCCAGAACAGCGATGGATGAATGGACAGTGTGCTGGAC 251
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
494 aPheAspValArgAlaGlyAlaIleAsnGlyThrValProLeuAsp 510
seq_name: sp_rodent:Q9EQR4

```

```

seq_documentation_block:
ID Q9EQR4 PRELIMINARY; PRT; 902 AA.
AC Q9EQR4;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE ENDOTHELIAL CHLORIDE CHANNEL.
GN CLC2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57/BL6;
RA Elble R.C., Pauli B.;
RT "A murine endothelial homolog of Lu-ECAM-1 associated with lung
melanoma metastasis."
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF115852; AAG47626.1;
DR InterPro: IPR003961; FN.III.
DR InterPro: IPR002035; VWFA.
DR SMART: SM00060; FN3; 1.
DR SMART: SM00327; VWFA; 1.
DR PROSITE: PS0234; VWFA; 1.
SQ SEQUENCE 902 AA; 99855 MW; F2BA57DAA5D4AAAD CRC64;

```

```

alignment_scores:
Quality: 235.00 Length: 83
Ratio: 3.264 Gaps: 1
Percent Similarity: 86.747 Percent Identity: 59.036

alignment_block:
US-09-049-696-8 x Q9EQR4
Align seg 1/1 to: Q9EQR4 from: 1 to: 902

```

```

3 CAAAGTGGTCCCATCATCCACAGTCGGCTTTGGGGCCCTCTCGAGCTCA 52
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
430 ArgSerGlyAlaIleIleHisThrIleAlaLeuGlyProSerAlaAlaAr 446
53 AGAAGTAGAGGAGCTGCCAAATGACAGAGGTTTACACACATATGCTT 102
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
446 gLuLeuGluThrLeuSerAspMetThrGlyGlyLeuArgPheThrAlaA 463
103 CAGATCAAGTTCAGAACAAATGGCTCATTCATGCTTTGGGGCCCTTCA 152
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
463 snlyHisVal.....SerSerLeuIleAspAlaPheSerArgIleSer 477

```

```

153 TCAGGAATGGAGCTGTCTTCAGCGCTCCATCCAGCTTGAGAGTAAGGG 202
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
478 SerThrSerGlySerValSerGlnGlnAlaLeuGlnLeuGluSerIysAl 494
203 ATTAACCCCTCCAGAACAGCGATGGATGAATGGACAGTGTGCTGGAC 251
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
494 aPheAsnValArgAlaGlyAlaIleAsnSerThrValProValAsp 510
seq_name: sp_human:Q9NXP1
seq_documentation_block:
ID Q9NXP1 PRELIMINARY; PRT; 469 AA.
AC Q9NXP1;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE CDNA FLJ20131 FIS, CLONE COL06357.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "N3DO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK000138; BAA90969.1;
SQ SEQUENCE 469 AA; 50694 MW; 6765986C886B3E56 CRC64;

```

```

alignment_scores:
Quality: 222.00 Length: 62
Ratio: 3.964 Gaps: 0
Percent Similarity: 90.323 Percent Identity: 67.742

alignment_block:
US-09-049-696-8 x Q9NXP1
Align seg 1/1 to: Q9NXP1 from: 1 to: 469

```

```

66 CTGTCCAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTCA 115
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MetSerLysIleThrGlySerHisPheThrValSerAspGluAlaCl 17
116 GAACAATGGCTCATTCATGCTTTGGGGCCCTTCATCAGGAATGGAG 165
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17 nasnasGlyLeuIleAspAlaPheGlyAlaLeuThrSerGlyAsnThrA 34
166 CTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAAACCTCCAG 215
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 spLeuSerGlnLysSerLeuGlnLeuGluSerLysGlyLeuThrLeuAsn 50
216 AACAGCAGTGGATGAATGGACAGCAGTGTGCTGGAC 251
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 SerAsnAlaTrpMetAsnAspThrValIleIleAsp 62
seq_name: sp_mammal:O18742

```

```

seq_documentation_block:
ID O18742 PRELIMINARY; PRT; 794 AA.
AC O18742;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OX NCBI_TaxID=9913;

```



```
seq_documentation_block:
```

DE NADP-DEPENDENT GLUTAMAT

DE NADP-DEPENDENT GLUTAMATE DEHYDROGENASE (EC 1.4.1.4) (FRAGMENT)

GN Giardia lamblia (Giardia intestinalis). Giardia.
 OS Eukaryota; Diplomonadida; Hexamitidae; Giardia.
 OC NCBI_TaxID=5741;
 OX [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=AD-135; PubMed=9481769;
 RX MEDLINE=98142768; PubMed=9481769;
 RA Monis P.T., Andrews R.H., Mayrhofer G., Mackrill J., Kulda J.,
 RA Isaac-Renton J.L., Ey P.L.;
 RT "Novel lineages of Giardia intestinalis identified by genetic analysis
 of organisms isolated from dogs in Australia.";
 RL Parasitology 116:0-0(0).
 DR EMBL; U60982; AAC17933.1;
 DR HSSP; P24295; LAUP.
 DR InterPro; IPR001625; GLFV_dehydrog.
 DR Pfam; PF00208; GLFV_dehydrog; 1.
 DR PRINTS; PRO0082; GLFDHGRNASE.
 DR PROSITE; PS00074; GLFV_DEHYDROGENASE; 1.
 KW Oxidoreductase.
 FT NON_TER 1
 FT NON_TER 230
 SQ SEQUENCE 230 AA; 25368 MW; FFE99E21C8561112 CRC64;

alignment_scores:
 Quality: 85.00 Length: 83
 Ratio: 1.771 Gaps: 3
 Percent Similarity: 57.831 Percent Identity: 33.735

alignment_block:
 US-09-049-696-8 x 077441 ..

Align seg 1/1 to: 077441 from: 1 to: 230

```

9 GGTGCCATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAACT 58
|||||:|||||: ||| ||| :|||: ||
142 GlySerLeuIleArgProGluAlaThrGlyTyrGlyAlaValTyrPheLe 158
|||||:|||||: ||| ||| :|||: ||
59 AGAGGAGCTGCTCAAAATGACAGGAGGTTTACAGACATATGCTTCAGATC 108
|||||:|||||: ||| ||| :|||: ||
158 uGluGluMetCysLys..... 163
109 AAGTTTCAGAACAAATGCCTCATATGATGCTTTGGGGCCCTTCATCAGGA 158
|||||:|||||: ||| ||| :|||: ||
164 .....AspAsnAsnThrIleIleArgGlyLysAsnValLeuLeuSerGly 178
159 AATGGAGCTGCTCTCAG.....CGCTCCATCCAGCTTGAGAG 196
|||||:|||||: ||| ||| :|||: ||
179 SerGlyAsnValAlaGlnPheAlaCysGluLysLeuIleGlnLeuGlyAl 195
|||||:|||||: ||| ||| :|||: ||
197 TAAGGGATTAAACCTCCAGACAGCCAGTGATGATGATGATGATGATGATC 245
|||||:|||||: ||| ||| :|||: ||
195 aLysValLeuThrPheSerAspSer.....AsnGlyThrIleVal 208
|||||:|||||: ||| ||| :|||: ||

```

seq_name: sp_invertebrate:097449

seq_documentation_block:
 ID 097449 PRELIMINARY; PRT: 230 AA.
 AC 097449;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE NADP-DEPENDENT GLUTAMATE DEHYDROGENASE (EC 1.4.1.4) (FRAGMENT).
 GN GDH.
 OS Giardia lamblia (Giardia intestinalis).
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.
 OX NCBI_TaxID=5741;
 OX [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=AD-141;
 RX MEDLINE=98142768; PubMed=9481769;
 RA Monis P.T., Andrews R.H., Mayrhofer G., Mackrill J., Kulda J.,

RA Isaac-Renton J.L., Ey P.L.;
 RT "Novel lineages of Giardia intestinalis identified by genetic analysis
 of organisms isolated from dogs in Australia.";
 RL Parasitology 116:0-0(0).
 DR EMBL; U60984; AAC17935.1;
 DR HSSP; P24295; LAUP.
 DR InterPro; IPR001625; GLFV_dehydrog.
 DR Pfam; PF00208; GLFV_dehydrog; 1.
 DR PRINTS; PRO0082; GLFDHGRNASE.
 DR PROSITE; PS00074; GLFV_DEHYDROGENASE; 1.
 KW Oxidoreductase.
 FT NON_TER 1
 FT NON_TER 230
 SQ SEQUENCE 230 AA; 25379 MW; 75913401C6881FF2 CRC64;

alignment_scores:
 Quality: 85.00 Length: 83
 Ratio: 1.771 Gaps: 3
 Percent Similarity: 57.831 Percent Identity: 33.735

alignment_block:
 US-09-049-696-8 x 097449 ..

Align seg 1/1 to: 097449 from: 1 to: 230

```

9 GGTGCCATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAACT 58
|||||:|||||: ||| ||| :|||: ||
142 GlySerLeuIleArgProGluAlaThrGlyTyrGlyAlaValTyrPheLe 158
|||||:|||||: ||| ||| :|||: ||
59 AGAGGAGCTGCTCAAAATGACAGGAGGTTTACAGACATATGCTTCAGATC 108
|||||:|||||: ||| ||| :|||: ||
158 uGluGluMetCysLys..... 163
109 AAGTTTCAGAACAAATGCCTCATATGCTTTGGGGCCCTTCATCAGGA 158
|||||:|||||: ||| ||| :|||: ||
164 .....AspAsnAsnThrIleIleArgGlyLysAsnValLeuLeuSerGly 178
159 AATGGAGCTGCTCTCAG.....CGCTCCATCCAGCTTGAGAG 196
|||||:|||||: ||| ||| :|||: ||
179 SerGlyAsnValAlaGlnPheAlaCysGluLysLeuIleGlnLeuGlyAl 195
|||||:|||||: ||| ||| :|||: ||
197 TAAGGGATTAAACCTCCAGACAGCCAGTGATGATGATGATGATGATGATC 245
|||||:|||||: ||| ||| :|||: ||
195 aLysValLeuThrPheSerAspSer.....AsnGlyThrIleVal 208
|||||:|||||: ||| ||| :|||: ||

```

seq_name: sp_invertebrate:097449

seq_documentation_block:
 ID 097449 PRELIMINARY; PRT: 230 AA.
 AC 097449;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE NADP-DEPENDENT GLUTAMATE DEHYDROGENASE (EC 1.4.1.4) (FRAGMENT).
 GN GDH.
 OS Giardia lamblia (Giardia intestinalis).
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.
 OX NCBI_TaxID=5741;
 OX [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=AD-147, AND AD-137;
 RX MEDLINE=98142768; PubMed=9481769;
 RA Monis P.T., Andrews R.H., Mayrhofer G., Mackrill J., Kulda J.,
 RA Isaac-Renton J.L., Ey P.L.;
 RT "Novel lineages of Giardia intestinalis identified by genetic analysis
 of organisms isolated from dogs in Australia.";
 RL Parasitology 116:0-0(0).
 DR EMBL; U60985; AAC17936.1;
 DR EMBL; U60983; AAC17934.1;
 DR HSSP; P24295; LAUP.
 DR InterPro; IPR001625; GLFV_dehydrog.
 DR Pfam; PF00208; GLFV_dehydrog; 1.


```

Quality: 83.00      Length: 83
Ratio: 1.729      Gaps: 3
Percent Similarity: 57.831      Percent Identity: 32.530

alignment_block:
US-09-049-696-8 x Q9XT8

```

[illegible]

138 UGIUGLUMETCYSLS.....

[illegible]

197 TAAGGATTACCCCTCCACAAGCAGTCGGATGAATGGCACACTGATC 245
||||| ||||| :||:::||| | ||||| ::|||
195 aLysValLeuThrPheSerAspSer AsnGlyThrIleVal 208

$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) = \frac{\partial L}{\partial x}$

100

100

1. *Chlorophyll a* (Chl *a*)

—

1. $\frac{1}{2}$ 2. $\frac{1}{2}$ 3. $\frac{1}{2}$ 4. $\frac{1}{2}$ 5. $\frac{1}{2}$ 6. $\frac{1}{2}$ 7. $\frac{1}{2}$ 8. $\frac{1}{2}$ 9. $\frac{1}{2}$ 10. $\frac{1}{2}$ 11. $\frac{1}{2}$ 12. $\frac{1}{2}$ 13. $\frac{1}{2}$ 14. $\frac{1}{2}$ 15. $\frac{1}{2}$ 16. $\frac{1}{2}$ 17. $\frac{1}{2}$ 18. $\frac{1}{2}$ 19. $\frac{1}{2}$ 20. $\frac{1}{2}$ 21. $\frac{1}{2}$ 22. $\frac{1}{2}$ 23. $\frac{1}{2}$ 24. $\frac{1}{2}$ 25. $\frac{1}{2}$ 26. $\frac{1}{2}$ 27. $\frac{1}{2}$ 28. $\frac{1}{2}$ 29. $\frac{1}{2}$ 30. $\frac{1}{2}$ 31. $\frac{1}{2}$ 32. $\frac{1}{2}$ 33. $\frac{1}{2}$ 34. $\frac{1}{2}$ 35. $\frac{1}{2}$ 36. $\frac{1}{2}$ 37. $\frac{1}{2}$ 38. $\frac{1}{2}$ 39. $\frac{1}{2}$ 40. $\frac{1}{2}$ 41. $\frac{1}{2}$ 42. $\frac{1}{2}$ 43. $\frac{1}{2}$ 44. $\frac{1}{2}$ 45. $\frac{1}{2}$ 46. $\frac{1}{2}$ 47. $\frac{1}{2}$ 48. $\frac{1}{2}$ 49. $\frac{1}{2}$ 50. $\frac{1}{2}$ 51. $\frac{1}{2}$ 52. $\frac{1}{2}$ 53. $\frac{1}{2}$ 54. $\frac{1}{2}$ 55. $\frac{1}{2}$ 56. $\frac{1}{2}$ 57. $\frac{1}{2}$ 58. $\frac{1}{2}$ 59. $\frac{1}{2}$ 60. $\frac{1}{2}$ 61. $\frac{1}{2}$ 62. $\frac{1}{2}$ 63. $\frac{1}{2}$ 64. $\frac{1}{2}$ 65. $\frac{1}{2}$ 66. $\frac{1}{2}$ 67. $\frac{1}{2}$ 68. $\frac{1}{2}$ 69. $\frac{1}{2}$ 70. $\frac{1}{2}$ 71. $\frac{1}{2}$ 72. $\frac{1}{2}$ 73. $\frac{1}{2}$ 74. $\frac{1}{2}$ 75. $\frac{1}{2}$ 76. $\frac{1}{2}$ 77. $\frac{1}{2}$ 78. $\frac{1}{2}$ 79. $\frac{1}{2}$ 80. $\frac{1}{2}$ 81. $\frac{1}{2}$ 82. $\frac{1}{2}$ 83. $\frac{1}{2}$ 84. $\frac{1}{2}$ 85. $\frac{1}{2}$ 86. $\frac{1}{2}$ 87. $\frac{1}{2}$ 88. $\frac{1}{2}$ 89. $\frac{1}{2}$ 90. $\frac{1}{2}$ 91. $\frac{1}{2}$ 92. $\frac{1}{2}$ 93. $\frac{1}{2}$ 94. $\frac{1}{2}$ 95. $\frac{1}{2}$ 96. $\frac{1}{2}$ 97. $\frac{1}{2}$ 98. $\frac{1}{2}$ 99. $\frac{1}{2}$ 100. $\frac{1}{2}$

Figure 1. The effect of the concentration of the *Agrobacterium* suspension on the transformation efficiency of *Agrobacterium* strains.

100

us-09-049-696-8:rsp

This Page Blank (uspto)

OM of: US-09-049-696-8 to: SwissProt_39:* out_format : pfs

Date: Mar 30, 2002 2:52 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frame+ntp.model -DEV=xlp
-O=/cgn2_1/USP00_pool/US09049696/funat_28032002_145238_2085/app_query.fasta_1.12579
-DB=SwissProt_39 -QFMT=fastan -SUFFIX=rs -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1_165 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-8

Query length: 253

Database: SwissProt_39*

Database sequences: 100059

Database length: 36664827

Search time (sec): 306.030000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
SwissProt_39:ECIC_BOVIN +	239.00	472.55	1.3e-19	903	! P54281 bos taurus (bovine).
SwissProt_39:DH4_GIALA +	82.00	154.42	0.1401	449	! P28724 giardia lamblia (giar
SwissProt_39:ATIA_MOUSE +	73.50	129.26	1.38	1149	! P70704 mus musculus (mouse)
SwissProt_39:ATIA_HUMAN +	72.50	127.09	1.80	1164	! Q9Y290 homo sapiens (human)
SwissProt_39:TMS2_HUMAN +	69.50	127.89	3.84	492	! O15393 homo sapiens (human)
SwissProt_39:PTRA_SCOLI +	69.50	122.45	3.94	962	! P05458 escherichia coli. pro
SwissProt_39:IGA_NEIGO +	69.00	117.64	4.59	1532	! P09790 neisseria gonorrhoeae
SwissProt_39:PGK_MYCPN +	68.50	127.33	4.96	409	! P78018 mycoplasma pneumoniae
SwissProt_39:ATIA_BOVIN +	67.50	116.88	6.74	1149	! Q29449 bos taurus (bovine)
SwissProt_39:ATPO_ARATH +	66.50	121.46	8.49	507	! P92549 arabidopsis thaliana
SwissProt_39:ATPO_BRANA +	66.50	121.46	8.49	507	! P22201 brassica napus (rape)
SwissProt_39:ATPO_RAPSA +	66.50	121.46	8.49	507	! P23413 rapanus sativus (rad
SwissProt_39:GUNS_BACAG +	66.00	122.35	9.60	400	! O85465 bacillus agaradhaeren
SwissProt_39:SM5B_MOUSE +	65.50	113.16	11.42	1093	! Q60519 canine adenovirus typ
SwissProt_39:FIBP_ADECT +	64.00	115.76	16.50	542	! Q65914 mus musculus (mouse)
SwissProt_39:V120_HSV1 +	64.00	109.85	17.00	1123	! P10221 herpes simplex virus
SwissProt_39:ATPO_BETVU +	63.50	115.29	18.78	506	! Q06735 beta vulgaris (sugar
SwissProt_39:ATPO_WAIZE +	63.50	115.26	18.78	508	! P05494 zea mays (maize). atp
SwissProt_39:ATPO_PHAVU +	63.50	115.26	18.78	508	! P24459 phaseolus vulgaris (k
SwissProt_39:ATPO_SOYBN +	63.50	115.26	18.78	508	! Q01915 glycine max (soybean)
SwissProt_39:ATPO_ORYSA +	63.50	115.24	18.78	509	! P15998 oryza sativa (rice).
SwissProt_39:ATPO_WHEAT +	63.50	115.24	18.78	509	! P12862 triticum aestivum (wh
SwissProt_39:ATPO_HELAN +	63.50	115.23	18.79	510	! P18260 helianthus annuus (co
SwissProt_39:ATPB_MOUSE +	63.00	107.61	22.17	1148	! P98200 mus musculus (mouse)
SwissProt_39:THS_METTL +	62.50	112.64	24.34	544	! Q93624 methanococcus thermo
SwissProt_39:CRK7_HUMAN +	62.50	104.46	25.57	1490	! Q9NY44 homo sapiens (human)
SwissProt_39:PGK_THETH +	62.00	114.33	27.63	389	! P09403 thermus aquaticus (sub
SwissProt_39:HUTH_STRCO +	62.00	112.10	27.94	512	! Q96W1 streptomyces coelicol
SwissProt_39:YMH8_YEAST +	62.00	112.08	27.94	513	! Q03104 saccharomyces cerevis
SwissProt_39:SPL_RAT +	62.00	108.60	28.44	788	! Q01714 rattus norvegicus (rat)
SwissProt_39:DYHC_EHENT +	62.00	94.75	30.49	4344	! P45444 emericella nidulans
SwissProt_39:ILVH_SALTY +	61.50	120.36	30.43	163	! P21622 salmonella typhimurii
SwissProt_39:YPVB_METTF +	61.50	112.10	31.73	451	! P29578 methanobacterium ther
SwissProt_39:MTRA_METTA +	61.00	115.92	35.34	248	! Q32867 methanopyrus kandleri
SwissProt_39:GSHI_MOUSE +	61.00	115.51	35.41	261	! P31315 mus musculus (mouse).
SwissProt_39:STX1_CAHEL +	61.00	114.40	35.61	299	! Q20024 caenorhabditis elegan
SwissProt_39:ITAL_MOUSE +	61.00	103.38	37.64	1163	! P24063 mus musculus (mouse)
SwissProt_39:CTIG_KLEPN +	60.50	113.29	40.66	302	! P45414 klebsiella pneumoniae
SwissProt_39:UPAR_RAT +	60.50	112.62	40.80	328	! P49616 rattus norvegicus (rat)
SwissProt_39:LAID_PSESP +	60.50	112.38	40.85	338	! Q00740 pseudomonas sp. (stra

alignment_scores:

Quality: 239.00 Length: 83

SwissProt_39:CYSS_DICDI + 60.50 112.23 40.88 344 ! P54640 dictyostelium disc
SwissProt_39:ATPO_PEA + 60.50 109.09 41.53 507 ! P05493 pisum sativum (garde
SwissProt_39:CIILA_KLEPN + 60.50 109.07 41.53 508 ! P45413 klebsiella pneumon
SwissProt_39:YAB4_MYCPN + 60.50 108.82 41.59 524 ! P75609 mycoplasma pneumon
SwissProt_39:FOUO_DROME + 60.00 106.92 47.67 583 ! P54360 drosophila melanog

seq_name: SwissProt_39:ECIC_BOVIN

seq_documentation_block:

ID ECIC_BOVIN STANDARD; PRT; 903 AA.
AC P54281;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN (CALCIUM-ACTIVATED CHLORIDE
DE CHANNEL).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Trachea;
RX MEDLINE=96125078; PubMed=8537359;
RA Cunningham S.A., Awayda M.S., Bubien J.K., Ismailov I.I.,
RA Arrate M.P., Berdies B.K., Benos D.J., Fuller C.M.;
RT "Cloning of an epithelial chloride channel from bovine trachea.";
RL J. Biol. Chem. 270:31016-31026(1995).
CC !- FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE
CC SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME;
CC MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND
CC TRANSEPITHELIAL TRANSPORT.
CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC !- TISSUE SPECIFICITY: TRACHEA.
CC !- PTM: PHOSPHORYLATED BY CAM-KINASE II.
CC !- SIMILARITY: CONTAINS 1 VWFA DOMAIN

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U36445; AAC48511.1; -
CC InterPro: IPR002035; VWFA.
CC SMART: SM00327; VWFA; 1.
CC PROSITE: PS0234; VWFA; 1.
CC Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
CC Calcium channel; Phosphorylation; Glycoprotein.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 331 351 POTENTIAL.
FT TRANSMEM 617 637 POTENTIAL.
FT TRANSMEM 883 903 POTENTIAL.
FT DOMAIN 308 476 VWFA.
FT CARBOHYD 75 75 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 688 688 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 811 811 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 816 816 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 842 842 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 857 857 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 903 AA; 100305 MW; 9742AB1590908AEC CRC64;

[illegible]

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U75321; AAB18627.1; ..
 DR MGD; MGI:1330848; Atpcl.
 DR InterPro; IPR001757; E1-E2_ATPase.
 DR InterPro; IPR001454; Hydrolase.
 DR Pfam; PF00122; E1-E2_ATPase; 1.
 DR Pfam; PF00702; Hydrolase; 1.
 DR PROSITE; PS00154; ATPASE_E1_E2; 1.
 KW Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
 KW Multigene family.

FT DOMAIN 1 65 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 66 86 POTENTIAL.
 FT DOMAIN 87 92 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 93 115 POTENTIAL.
 FT DOMAIN 116 297 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 298 319 POTENTIAL.
 FT DOMAIN 320 344 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 345 366 POTENTIAL.
 FT DOMAIN 367 842 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 843 863 POTENTIAL.
 FT DOMAIN 864 875 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 876 895 POTENTIAL.
 FT DOMAIN 896 925 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 926 947 POTENTIAL.
 FT DOMAIN 948 961 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 962 984 POTENTIAL.
 FT DOMAIN 985 990 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 991 1011 POTENTIAL.
 FT DOMAIN 1012 1029 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1030 1055 POTENTIAL.
 FT DOMAIN 1056 1149 CYTOPLASMIC (POTENTIAL).
 FT MOD_RES 409 409 PHOSPHORYLATION (BY SIMILARITY).
 FT METAL 786 786 MAGNESIUM (BY SIMILARITY).
 FT METAL 790 790 MAGNESIUM (BY SIMILARITY).
 FT NP_BIND 726 733 ATP (POTENTIAL).
 FT NP_BIND 1080 1087 ATP (POTENTIAL).
 SQ SEQUENCE 1149 AA; 129766 MW; 95F70B2B091FD31E CRC64;

alignment_scores:
 Quality: 73.50 Length: 69
 Ratio: 1.750 Gaps: 2
 Percent similarity: 60.870 Percent Identity: 26.087

alignment_block:

US-09-049-696-8 x ATIA_MOUSE ..

Align seg 1/1 to: ATIA_MOUSE from: 1 to: 1149

42 TCTCAGCTCAAGAACTAGAGAGCTGTCAAAATGACAGGAGTTTACA 91
 |||
 204 SerAspIleLysAspIleAspSerLeuMetArgIleSerGlyArgIleG 220
 92 GACATATGCTTCAGATCAAGTTCAGAAACAATGGCTCATTCGCTTTTG 141
 :
 220 ucys.....GlusPrAsnArgHisLeuTyrAspPheValG 233
 142 GGGCCCTTTCA...TCAGGAATGAGAGCTGTCTTCAGCGCTCCATCCAG 188
 ||
 233 LyAsnIleArgLeuAspGlyHisGlyThrValProLeuGlyAlaAspGln 249
 189 CTTGAGAGTAGGAGTTAACCTCCAGACAGCCAGTGGATGATGGCAC 238
 :
 250 IleLeuLeuArgGlyAlaGlnLeuArgAsnThrGlnTrpValHisGlyI 266
 239 AGTGATC 245
 |||...
 266 evalVal 268

seq_name: SwissProt_39:ATIA_HUMAN

seq_documentation_block:

ID ATIA_HUMAN STANDARD; PRT; 1164 AA.

AC Q9Y200;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE POTENTIAL PHOSPHOLIPID-TRANSPORTING ATPASE IA (EC 3.6.3.13)
 GN (CHROMAFFIN GRANULE ATPASE II).
 BE ATP8A1 OR ATP1A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN 1;
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=99216279; PubMed=10198212;
 RA Muro I., Halleck M.S., Schlegel R.A., Mattel M.-G., Williamson P.L.,
 RA Zaczowski A., Devaux P., Cartron J.-P., Colin Y.;
 RT "Cloning, expression, and chromosomal mapping of a human ATPase II
 RT gene, member of the third subfamily of P-type ATPases and orthologous
 RT to the presumed bovine and murine aminophospholipid translocase.";
 RL Biochem. Biophys. Res. Commun. 257:333-339(1999).
 RN 121;
 RP SEQUENCE OF 4-1164 FROM N.A. (LONG ISOFORM).
 RC TISSUE=Brain;
 RA Osada S., Nakanishi Y.;
 RT "cDNA cloning of human ATPaseII.";
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE TRANSPORT OF AMINOPHOSPHOLIPIDS
 CC FROM THE OUTER TO THE INNER LEAFLET OF VARIOUS MEMBRANES AND THE
 CC MAINTENANCE OF ASYMMETRIC DISTRIBUTION OF PHOSPHOLIPIDS, MAINLY IN
 CC SECRETORY VESICLES.
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CHROMAFFIN
 CC GRANULES (BY SIMILARITY).
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG ISOFORM (SHOWN HERE)
 CC AND A SHORT ISOFORM; ARE PRODUCED BY TISSUE SPECIFIC ALTERNATIVE
 CC SPLICING.
 CC -!- TISSUE SPECIFICITY: FOUND IN MOST ADULT TISSUES EXCEPT LIVER.
 CC TESTIS AND PLACENTA. MOST ABUNDANT IN HEART, BRAIN AND SKELETAL
 CC MUSCLE. ALSO DETECTED IN FETAL TISSUES. THE LONG ISOFORM IS ONLY
 CC DETECTED IN BRAIN, SKELETAL MUSCLE AND HEART AND IS THE MOST
 CC ABUNDANT FORM IN SKELETAL MUSCLE.
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2
 CC ATPASES). SUBFAMILY IV.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF067820; AAD34706.1; ..
 DR EMBL; AB013452; BAA77248.1; ..
 DR InterPro; IPR001757; E1-E2_ATPase.
 DR InterPro; IPR001454; Hydrolase.
 DR Pfam; PF00122; E1-E2_ATPase; 1.
 DR Pfam; PF00702; Hydrolase; 1.
 DR PRINTS; PR00119; CATATPASE.
 DR PROSITE; PS00154; ATPASE_E1_E2; 1.
 KW Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
 KW Alternative splicing; Multigene family.
 FT DOMAIN 1 65 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 66 86 POTENTIAL.
 FT DOMAIN 87 92 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 93 115 POTENTIAL.
 FT DOMAIN 116 297 CYTOPLASMIC (POTENTIAL).

seq_name	seq_start	seq_end	seq_type	seq_desc
TRANSMEM	298	319	POTENTIAL	
DOMAIN	320	344	EXTRACELLULAR (POTENTIAL)	
TRANSMEM	345	366	POTENTIAL	
DOMAIN	367	387	CYTOPLASMIC (POTENTIAL)	
TRANSMEM	858	878	POTENTIAL	
DOMAIN	879	890	EXTRACELLULAR (POTENTIAL)	
TRANSMEM	891	910	POTENTIAL	
DOMAIN	911	940	CYTOPLASMIC (POTENTIAL)	
TRANSMEM	941	962	POTENTIAL	
DOMAIN	963	976	EXTRACELLULAR (POTENTIAL)	
TRANSMEM	977	999	POTENTIAL	
DOMAIN	1000	1005	CYTOPLASMIC (POTENTIAL)	
TRANSMEM	1006	1026	POTENTIAL	
DOMAIN	1027	1044	EXTRACELLULAR (POTENTIAL)	
TRANSMEM	1045	1070	POTENTIAL	
DOMAIN	1071	1164	CYTOPLASMIC (POTENTIAL)	
MOD_RES	409	409	PHOSPHORYLATION (BY SIMILARITY)	
METAL	801	801	MAGNESIUM (BY SIMILARITY)	
METAL	805	805	MAGNESIUM (BY SIMILARITY)	
NP_BIND	741	748	ATP (POTENTIAL)	
NP_BIND	1095	1102	ATP (POTENTIAL)	
VARSPLIC	433	447	MISSING (IN SHORT ISOFORM)	
CONFLICT	364	364	E -> K (IN REF. 2)	
SEQUENCE	1164 AA	131369 MW	CE1EAF0206CD36F7 CRC64	

alignment_scores:

Quality:	72.50	Length:	69
Ratio:	1.726	Gaps:	2
Percent Similarity:	60.870	Percent Identity:	26.087

alignment_block:

US-09-049-696-8 x AT1A_HUMAN

Align seg 1/1 to: AT1A_HUMAN from: 1 to: 1164

```

42  TCTGCGAGCTCAAGAACTAGAGGAGCTCTCCAAATGACAGAGGTTTACA 91
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 SerAspIleLeuAspValSerLeuMetArgIleSerGlyArgIleG 220
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
92  GACATATGCTTCAGATCAAGTTCAGAACATGCGCTCATTTGCTTTG 141
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
220 uCys.....GluSerProAsnArgHisLeuTyAspPheValG 233
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
142 GGGCCCTTTCA...TCAGAAATGAGCTGTCTCTCAGCGCTCCATCAG 188
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
233 LyAsnIleArgLeuAspGlyHisGlyThrValProLeuGlyAlaAspGln 249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
189 CTTCAGATTAAGGATTAACCTCCAGAACCCAGTCGATGATGATGCGAC 238
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
250 IleLeuLeuArgGlyAlaGlnLeuArgAsnThrGlnTrpValHisGlyI 266
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
239 AGTGATC 245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
266 eValVal 268
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

seq_name: SwissProt_39:TMS2_HUMAN

seq_documentation_block:

ID	TMS2_HUMAN	STANDARD;	PRT;	492 AA.
AC	O15393;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-).			
GN	TMPSR2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
PN	[1]			
PR	SEQUENCE FROM N.A.			

Tue Apr 2 09:40:14 2002

us-09-049-696-8.isp

FT MUTAGEN 162 162 E->Q: 20% LOSS OF ACTIVITY.
FT MUTAGEN 169 169 E->Q: LOSS OF ACTIVITY AND OF 2N-BINDING.
FT MUTAGEN 204 204 E->Q: NO LOSS OF ACTIVITY.
FT CONFLICT 277 284 IITHVPA -> HVHSLRPW (IN REF. 4).
SQ SEQUENCE 962 AA; 107708 MW; 0558C68C2F1A0540 CRC64;

alignment_scores:
Quality: 69.50 Length: 89
Ratio: 1.479 Gaps: 3
Percent Similarity: 52.809 Percent Identity: 30.337
alignment_block:
US-09-049-696-8 x PTR_A_ECOLI ..
Align seg 1/1 to: PTR_A_ECOLI from: 1 to: 962

3 CAAGTGGTCCATCCACAGTCGCTTTGGGGCCCTCTGCAGTCA 52
|||||
274 GlnLysGlyLeuIleHisTyValProAlaLeuProArgLysValle 290
53 AGAACTAGAG.....GAGCTGTCCAAATGA 78
290 uArgValGluPheArgIleAspAsnSerAlaLysPheArgSerLysT 307
79 CAGAGGTTTACAGCATAT.....GCT 101
307 hrAspGluLeuIleThrTyLeuIleGlyAsnArgSerProGlyThrLeu 323
102 TCAGATCAAGTTTCAGAAATGCGCTCATGATGCTTTGGGGCCCTTTC 151
324 SerAspTrpLeuGlnLysGlnGlyLeuValGluGlyLeuSerAlaAsnSe 340
152 A.....TCAGAAATGAGCTGTCTCTCAGCGCTCCATCCAGC 189
340 rAspProIleValAsnGlyAsnSerGlyValLeuAlaIleSerAlaSerL 357
190 TTGAGAGTAAGGGATTA 206
357 euThrAspLysGlyLeu 362

seq_name: SwissProt_39:IGA_NEIGO

seq_documentation_block:
ID IGA_NEIGO STANDARD; PRT; 1532 AA.
AC P09790;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE IGA-SPECIFIC SERINE ENDOPEPTIDASE PRECURSOR (EC 3.4.21.72) (IGA
PROTEASE).
GN IGA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=MS11;
RX MEDLINE=871115823; PubMed=3027577;
RA Pohlner J., Halter R., Beyreuther K., Meyer T.F.;
RT "Gene structure and extracellular secretion of Neisseria gonorrhoeae
RT IGA protease";
RL Nature 325:458-462(1987).
RN [2]
RP ACTIVE SITE.
RX MEDLINE=90154052; PubMed=2105953;
RA Bachovchin W.W., Plaut A.G., Flentner G.R., Lynch M., Kettner C.A.;
RT "Inhibition of Iga1 proteinases from Neisseria gonorrhoeae and
RT Hemophilus influenzae by peptide prolyl boronic acids";
RL J. Biol. Chem. 265:3738-3743(1990).
CC -1- FUNCTION: THIS PROTEASE IS SPECIFIC FOR IMMUNOGLOBULIN A.
CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF IMMUNOGLOBULIN A MOLECULES AT
CC CERTAIN PRO-1-XAA BONDS IN THE HINGE REGION. NO SMALL MOLECULE

CC SUBSTRATES ARE KNOWN.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC -1- DOMAIN: THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X04835; CAA28538.1; ..
CC EIR; A26039; A26039.
CC MEROPS; S06.001; ..
CC InterPro; IPR000710; IGA_S6.
CC Pfam; PF02395; IGA1; 1.
CC PRINTS; PR00921; IGASERPTASE.
CC Hydrolase; Serine protease; Zymogen; Autocatalytic cleavage;
KW Transmembrane; Signal.
FT SIGNAL 1 27 IGA-SPECIFIC SERINE ENDOPEPTIDASE.
FT CHAIN 28 986 HELPER PEPTIDE.
FT PROPEP 987 1532 POTENTIAL.
FT ACT_SITE 278 278 CLEAVAGE (AUTO-).
FT SITE 986 987 CLEAVAGE (AUTO-).
FT SITE 1018 1019 CLEAVAGE (AUTO-).
FT SITE 1121 1122 CLEAVAGE (AUTO-).
SQ SEQUENCE 1532 AA; 168976 MW; 68FF4112BD22F40D CRC64;

alignment_scores:
Quality: 69.00 Length: 77
Ratio: 1.438 Gaps: 2
Percent Similarity: 62.338 Percent Identity: 28.571
alignment_block:
US-09-049-696-8 x IGA_NEIGO ..
Align seg 1/1 to: IGA_NEIGO from: 1 to: 1532

18 ATCCACACAGTCGCTTTGGGGCCCTCTCAGCTCAGACTAGAGGAGCT 67
|||||
1221 ValGluThrAlaGlyLeuGlnProArgAlaAlaGlnProArgThrGlnAl 1237
68 GTCCAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTTCA 117
|||||
1237 aAlaAlaGlnAlaAspAlaValSerThrAsnThr.....A 1249
118 ACAATGGCCTCATTTGATGCTTTTGGGGCCCTTTTCATCA.....CGA 158
|||||
1249 snSerAlaLeuSerAspAlaMetAlaSerThrGlnSerIleLeuLeuAsp 1265
159 AATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGATGAAGGGATTAC 208
|||||
1266 ThrGlyAlaTyrlLeuThrArgHisIleAlaGlnLysSerArgAlaAspAl 1282
209 CTTCCAGAACAGCCAGTGGATGAATGGCACA 239
1282 aGluLysAsnSerValTrpMetSerAsnThr 1292

seq_name: SwissProt_39:PGK_MYCPN

seq_documentation_block:
ID PGK_MYCPN STANDARD; PRT; 409 AA.
AC P78018;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHOSPHOGLYCERATE KINASE (EC 2.7.2.3).

GN PGK OR MPN429 OR MP412.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE = ADP +
CC 3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

DR EMBL: AE000040; AAB96060.1;
DR HSP; P36204; 1VPE.
DR InterPro: IPR001576; PGK.
DR Pfam: PF00162; PGK; 1.
DR PRINTS: PR00477; PHGLYCKINASE.
DR PROSITE: PS00111; GLYCYCKINASE.
KW Transferase; Kinase; Glycolysis; Complete proteome.
SQ SEQUENCE 409 AA; 44212 MW; 6AF2301880398731 CRC64;

alignment_scores:
Quality: 68.50 Length: 48
Ratio: 2.076 Gaps: 1
Percent Similarity: 68.750 Percent Identity: 35.417

alignment_block:

US-09-049-696-8 x PGK_MYCPN ..

Align seg 1/1 to: PGK_MYCPN from: 1 to: 409

24 ACAGTCGCTTGGGCGCTTCGACCTCAAGCACTAGAGGAGCTGCCAA 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
346 ThrLeuAlaIleGlyLysSerAlaAlaAlaGlnGlnAlaPheSerVa 362
74 AATGACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGACAATG 123
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||
362 lIleGlyGlyGlyAspSerAlaAlaAlaLysGlnLeuGln..... 376
124 GCCTCATGTATGCTTTGGGGCGCTTCATCAGGAATGGAGCT 167
::: ||| |||:|||||:|||||:|||||:|||||:|||||:|||||
377 ..lIleAlaAspGlnPheSerPheIleSerThrGlyGlyGlyAla 390

seq_name: SwissProt_39.ATIA_BOVIN

seq_documentation_block:

ID ATIA_BOVIN STANDARD; PRT; 1149 AA.
AC Q29449;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE POTENTIAL PHOSPHOLIPID-TRANSPORTING ATPASE IA (EC 3.6.3.13)
DE (CHROMAFFIN GRANULE ATPASE II).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 502-678 AND 1144-1149.
RC TISSUE=Adrenal medulla;
RX MEDLINE=96234315; PubMed=8633245;
RA Tang X., Halleck M.S., Schlegel R.A., Williamson P.L.;
RT "A subfamily of P-type ATPases with aminophospholipid transporting
RT activity";
RL Science 272:1495-1497(1996).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE TRANSPORT OF AMINOPHOSPHOLIPIDS
CC FROM THE OUTER TO THE INNER LEAFLET OF VARIOUS MEMBRANES AND THE
CC MAINTENANCE OF ASYMMETRIC DISTRIBUTION OF PHOSPHOLIPIDS, MAINLY IN
CC SECRETORY VESICLES.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CHROMAFFIN
CC GRANULES.
CC -1- TISSUE SPECIFICITY: KIDNEY.
CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2
CC ATPASES); SUBFAMILY IV.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

DR EMBL: U51100; AAD03352.1;
DR InterPro: IPR001757; E1-E2_ATPase.
DR InterPro: IPR001454; Hydrolase.
DR Pfam: PF00122; E1-E2_ATPase; 1.
DR Pfam: PF00702; Hydrolase; 1.
DR PROSITE: PS00154; ATPASE_E1_E2; 1.
KW Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
KW Multigene family.

ET DOMAIN 1 65 CYTOPLASMIC (POTENTIAL).
ET TRANSMEM 66 86 POTENTIAL.
ET DOMAIN 87 92 EXTRACELLULAR (POTENTIAL).
ET TRANSMEM 93 115 POTENTIAL.
ET DOMAIN 116 297 CYTOPLASMIC (POTENTIAL).
ET TRANSMEM 298 319 POTENTIAL.
ET DOMAIN 320 344 EXTRACELLULAR (POTENTIAL).
ET TRANSMEM 345 366 POTENTIAL.
ET DOMAIN 367 842 CYTOPLASMIC (POTENTIAL).
ET TRANSMEM 843 863 POTENTIAL.
ET DOMAIN 864 875 EXTRACELLULAR (POTENTIAL).
ET TRANSMEM 876 895 POTENTIAL.
ET DOMAIN 896 925 CYTOPLASMIC (POTENTIAL).
ET TRANSMEM 926 947 POTENTIAL.
ET DOMAIN 948 961 EXTRACELLULAR (POTENTIAL).
ET TRANSMEM 962 984 POTENTIAL.
ET DOMAIN 985 990 CYTOPLASMIC (POTENTIAL).
ET TRANSMEM 991 1011 POTENTIAL.
ET DOMAIN 1012 1029 EXTRACELLULAR (POTENTIAL).
ET TRANSMEM 1030 1055 POTENTIAL.
ET DOMAIN 1056 1149 CYTOPLASMIC (POTENTIAL).
ET MOD_RES 409 409 PHOSPHORYLATION (BY SIMILARITY).
ET METAL 786 786 MAGNESIUM (BY SIMILARITY).
ET METAL 790 790 MAGNESIUM (BY SIMILARITY).
ET NP_BIND 726 733 ATP (POTENTIAL).
ET NP_BIND 1080 1087 ATP (POTENTIAL).
SQ SEQUENCE 1149 AA; 130025 MW; 0BE71C958E8C5BE1 CRC64;

alignment_scores:

Quality: 67.50 Length: 69
Ratio: 1.607 Gaps: 2
Percent Similarity: 60.870 Percent Identity: 24.638

alignment_block:

"Sequence and transcript analysis of the Nco2.5 Ogura-specific fragment correlated with cytoplasmic male sterility in Brassica cybrids.";
 RL Mol. Gen. Genet. 235:340-348(1992).
 CC !- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON GRADIENT ACROSS THE MEMBRANE. THE ALPHA CHAIN IS A REGULATORY SUBUNIT.
 CC !- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0) HAS THREE MAIN SUBUNITS: A, B AND C.
 CC !- SUBCELLULAR LOCATION: MITOCHONDRIAL.
 CC !- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.

 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 DR EMBL; X56008; CAA39483.1; -;
 DR EMBL; J12627; CAA78274.1; -;
 DR PIR; S13382; PWRPA.
 DR HSSP; P19483; ICOW.
 DR InterPro; IPR000790; ATPase_A.C.
 DR Pfam; PF00422; ATP-synt_A-c; 1.
 DR Pfam; PF00006; ATP-synt_ab; 1.
 DR ProDom; PD001099; ATPase_A.C; 1.
 DR PROSITE; PS00152; ATPASE_ALPHA.BETA; 1.
 KW ATP synthesis; CF(1); Hydrogen ion transport;
 KW Hydroxylase; ATP-binding; Mitochondrion.
 FT NP_BIND 171 178 ATP (BY SIMILARITY).
 FT ACT_SITE 373 373 BY SIMILARITY.
 SQ SEQUENCE 507 AA; 55142 MW; 3E067B5F9ADBCBEA CRC64;

alignment_scores:
 Quality: 66.50 Length: 68
 Ratio: 1.446 Gaps: 1
 Percent Similarity: 67.647 Percent Identity: 26.471

alignment_block:
 US-09-049-696-8 x ATP0_BRANA ..

Align seq 1/1 to: ATP0_BRANA from: 1 to: 507

```

27  GTGCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGAGCTGTCCAAAT 76
    |||:::|||||:::|||||:::  ::  ::::|  |||
76  ValValPheGlyGlyAspThrAlaIleLysGluGlyAspLeuValLysAr 92
    |||:::|||||:::|||||:::  ::  ::::|  |||
77  GACAGGAGGTTTACAGACATATGCTCAGATCAAGTTTCAGAACAAATGCC 126
    |||:::|||||:::|||||:::  ::  ::::|  |||
92  gThrGlySerIleValAspValProAlaGlyLysAlaMetLeuGlyArgV 109
    |||:::|||||:::|||||:::  ::  ::::|  |||
127 TCATTGATGCTTTGGGGCCCTTTCATCAGAAATGAGCTGTCTCT... 173
    ::::|  |||:::|||||:::|||||:::  ::  ::::|  |||
109 alValAspAlaMetGlyValProIleAspGlyArgGlyAlaLeuSerAsp 125
    |||:::|||||:::|||||:::  ::  ::::|  |||
174 .....CAGGCTCCATCCAGCTTGCAGCTAAGGATTACCCCTCCAGAA 217
    |||:::|||||:::|||||:::  ::  ::::|  |||
126 HisGluClnArgArgValGluValLysAlaProGlyIleLeuGluArgLy 142
    |||:::|||||:::|||||:::  ::  ::::|  |||
218 CAGC 221
142 sSer 143
```

seq_name: SwissProt_39:ATP0_RAPSA

seq_documentation_block:

ID ATP0_RAPSA STANDARD; PRT; 507 AA.

```

E23413;
01-NOV-1991 (Rel. 20, Created)
01-NOV-1991 (Rel. 20, Last sequence update)
30-MAY-2000 (Rel. 39, Last annotation update)
DE ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL (EC 3.6.1.34).
CN ATP.
OS Raphanus sativus (Radish), and Brassica campestris (Field mustard).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
  eurosids II; Brassicales; Brassicaceae; Raphanus.
OX NCBI_TaxID=3726, 3711;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES= R. sativus;
RX MEDLINE=91346710; PubMed=2151720;
RA Makaroff C.A., Apel I.J., Palmer J.D.;
  "Characterization of radish mitochondrial atpA: influence of nuclear
  background on transcription of atpA-associated sequences and
  relationship with male sterility.";
  Plant Mol. Biol. 15:735-746(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES= B. campestris; STRAIN=CV. PURPLE TOP GOLDEN GLOBE;
RA Hammett L.A., Nielsen B.L.;
  "DNA sequence of the Brassica campestris mitochondrial atpA gene.";
  (In) Plant Gene Register PGR98-186.
CC !- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
  GRADIENT ACROSS THE MEMBRANE. THE ALPHA CHAIN IS A REGULATORY
  SUBUNIT.
CC !- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
  CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
  SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
  HAS THREE MAIN SUBUNITS: A, B AND C.
CC !- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC !- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
  -----
  CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
  -----
  DR EMBL; AF076166; AAC78472.1; -;
  DR PIR; S12309; S12309.
  DR HSSP; P19483; ICOW.
  DR InterPro; IPR000790; ATPase_A.C.
  DR InterPro; IPR000194; ATPase_alpha_beta.
  DR Pfam; PF00422; ATP-synt_A-c; 1.
  DR Pfam; PF00006; ATP-synt_ab; 1.
  DR ProDom; PD001099; ATPase_A.C; 1.
  DR PROSITE; PS00152; ATPASE_ALPHA.BETA; 1.
  KW ATP synthesis; CF(1); Hydrogen ion transport;
  KW Hydroxylase; ATP-binding; Mitochondrion.
  FT NP_BIND 171 178 ATP (BY SIMILARITY).
  FT ACT_SITE 373 373 BY SIMILARITY.
  SQ SEQUENCE 507 AA; 55057 MW; 3E06628F9ADBCBEA CRC64;
  -----
  alignment_scores:
    Quality: 66.50 Length: 68
    Ratio: 1.446 Gaps: 1
    Percent Similarity: 67.647 Percent Identity: 26.471
  alignment_block:
    US-09-049-696-8 x ATP0_RAPSA ..
  Align, Seq 1/1 to: ATP0_RAPSA from: 1 to: 507
  -----
  27  GTGCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGAGCTGTCCAAAT 76
      |||:::|||||:::|||||:::  ::  ::::|  |||
      |||:::|||||:::|||||:::  ::  ::::|  |||
  27  GTGCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGAGCTGTCCAAAT 76
      |||:::|||||:::|||||:::  ::  ::::|  |||
      |||:::|||||:::|||||:::  ::  ::::|  |||
```

us-09-049-696-8.rsp

Tue Apr 2 09:40:14 2002

```

76 ValValPheGlyGlyAspThrAlaIleIleGlyGluGlyAspLeuValLysAr 92
DR InterPro: IPR003610; Chitin_bind3.
DR InterPro: IPR001547; Glyco_hydro_F5.
DR Pfam: PF001150; cellulase; 1.
DR SMART: SM00495; ChtBD3; 1.
DR PROSITE: PS00659; GLYCOSYL-HYDROL_F5; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; 3D-structure.
FT ACT_SITE 165 165 PROTON DONOR.
FT ACT_SITE 254 254 NUCLEOPHILE.
FT SEQUENCE 400 AA; 44702 MW; 3f9c66f9b9c36fff CRC64;
SQ

alignment_scores:
Quality: 66.00 Length: 46
Ratio: 2.062 Gaps: 2
Percent Similarity: 69.565 Percent Identity: 34.783

alignment_block:
3S:09-049-696-8 x GUN5_BACAG ..
Align seg 1/1 to: GUN5_BACAG from: 1 to: 400

93 ACATATGCTTCAGATCAAGTTCCAGAAATGCGCTCATGATGCTTTGG 142
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
23 ThrThrAlaAlaasp.....AsnAspSerValValGluGluHisGl 36

143.GGCCCTTTCATCAGGAATGGA...GCTGTCTCTCAGCGCTCCATCCAGC 189
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
36 yGlnLeuSerIleSerAsnGlyGluLeuValAsnGluArgGlyGluGlnV 53

190 TTGAGAGTAAGGATTAAACCTCCAGAACACAGCCAGTGG 227
||||| |||||| |||||| |||||| |||||| |||||| |||||| ||||||
53 aGlnLeuLysGlyMetSerSerHisGlyLeuGlnTrp 65

seq_name: SwissProt_39:SM5B_MOUSE

seq_documentation_block:
ID: SM5B_MOUSE STANDARD; PRT: 1093 AA.
AC: Q60519;
DT: 30-MAY-2000 (Rel. 39, Created)
DI: 30-MAY-2000 (Rel. 39, Last sequence update)
ET: 20-AUG-2001 (Rel. 40, Last annotation update)
PE: SEMAPHORIN 5B PRECURSOR (SEMAPHORIN G) (SEMA G).
GN: SEMA5B OR SEMAG OR SEMG.
OS: Mus musculus (Mouse).
OC: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC: Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX: NCBI_TaxID=10090;
RN: [1];
RP: SEQUENCE FROM N.A.
RC: STRAIN=NMRI;
RX: MEDLINE=96414430; PubMed=8817451;
RA: Adams R.H., Betz H., Pueschel A.W.;
RT: "A novel class of murine semaphorins with homology to thrombospondin
is differentially expressed during early embryogenesis.";
RL: Mech. Dev. 57:33-45(1996).
CC: -!- FUNCTION: MAY ACT AS POSITIVE AXONAL GUIDANCE CUES.
CC: -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC: -!- TISSUE SPECIFICITY: IN ADULT, ONLY DETECTED IN BRAIN.
CC: -!- DEVELOPMENTAL STAGE: DIFFERENTIALLY EXPRESSED IN EMBRYONIC AND
ADULT TISSUES. ITS ABUNDANCE DECREASES FROM E10 TO BIRTH.
CC: -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC: -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC: -!- SIMILARITY: CONTAINS 7 TSP TYPE-1 DOMAINS.
CC: This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```


OM of: US-09-049-696-8 to: PIR_68:* out_format : pfs

Date: Mar 30, 2002 2:26 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frame+2np.model -DEV=xlp
-O/cgn2_1/USPRO_spool/US09049696/runat_28032002_145238_2031/app_query.fasta_1.12579
-DB=PIR_68 -QFMT=fastan -SUFFIX=rpr -GAPOB=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOB=4.500
-GAPEXT=0.050 -GAPOB=10.000 -GAPEXT=0.500 -GAPEXT=6.000
-GAPEXT=7.000 -GAPOB=10.000 -GAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09049696_eCGN1_1_367 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-049-696-8

Query length: 253

Database: PIR_68:

Database sequences: 219241

Database length: 76174552

Search time (sec): 470.790000

score.list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
pir2:T02205	+	62.00	432.66	4.6e-17	905 Lu-ECAM-1 protein - bovine
pir1:A42489	+	82.00	156.27	0.2293	449 glutamate dehydrogenase (NADP+)
pir2:T28698	+	76.50	132.02	0.9983	2314 hypothetical protein - Streptococcus
pir2:C81169	+	74.00	128.64	1.96	1815 IgA-specific metalloendopeptidase
pir2:T70133	+	73.50	141.65	2.28	295 ATP-binding protein (vlyH-1) h
pir2:T30869	+	73.50	131.13	2.25	1149 probable adenosinetriphosphat
pir1:SNCEPI	+	69.50	124.09	6.84	962 ptilinysin (EC 3.4.24.55) precu
pir2:H85933	+	69.50	124.09	6.84	962 proteinase III [imported] - Esc
pir2:T74325	+	69.00	133.46	7.68	250 hypothetical protein slr0431 -
pir2:S31831	+	69.00	130.43	7.65	370 IgA-specific metalloendopeptida
pir2:T40431	+	69.00	130.43	7.65	370 IgA-specific metalloendopeptida
pir2:T26039	+	69.00	119.44	7.57	1532 IgA-specific metalloendopeptida
pir2:T73738	+	68.50	128.60	8.75	409 phosphoglycerate kinase (EC 2.7
pir2:T82750	+	68.00	111.44	9.86	3282 hemagglutinin-like secreted pr
pir2:T82589	+	68.00	111.07	9.85	3442 hemagglutinin-like secreted pr
pir2:T82519	+	68.00	111.04	9.85	3455 hemagglutinin-like secreted pr
pir2:T70951	+	67.50	119.17	11.38	1055 hypothetical protein Rv3202c -
pir2:T18515	+	67.50	118.51	11.37	1149 adenosinetriphosphatase (EC 3.
pir2:T40432	+	67.00	127.25	13.14	324 IgA-specific metalloendopeptida
pir2:T38820	+	67.00	116.11	13.00	1367 probable calcium-transporting
pir2:T61314	+	67.00	115.09	12.98	1561 IgA-specific metalloendopeptid
pir2:T81937	+	67.00	114.10	12.97	1773 IgA-specific metalloendopeptid
pir1:PWRA	+	66.50	122.73	14.99	507 H+-transporting ATP synthase (H
pir2:T12309	+	66.50	122.73	14.99	507 H+-transporting ATP synthase (H
pir2:T86368	+	66.50	116.32	14.89	1161 protein F28C11.9 [imported] -
pir2:T06726	+	65.50	123.22	19.68	363 cysteine proteinase (EC 3.4.22.
pir2:T80619	+	64.50	123.18	25.82	278 homeotic protein cnox1 - hydrom
pir2:T84367	+	64.00	119.51	29.48	390 hypothetical protein Vng2168c [
pir2:T49219	+	64.00	116.96	29.40	542 fiber protein - canine adenovin
pir2:TJ5930	+	64.00	115.58	29.36	648 serine/arginine-rich protein-sh
pir1:WBEH7	+	64.00	111.32	29.24	1123 UL37 protein - human herpesvir
pir4:T29876	+	63.50	119.82	33.78	327 probable H+-transporting ATP sy
pir2:T53922	+	63.50	116.44	33.66	506 H+-transporting ATP synthase (H
pir2:T84608	+	63.50	116.44	33.66	506 H+-transporting ATP synthase (H
pir1:PWZMAM	+	63.50	116.41	33.66	508 H+-transporting ATP synthase (H
pir2:T26979	+	63.50	116.41	33.66	508 H+-transporting ATP synthase (H
pir2:T29792	+	63.50	116.40	33.66	509 H+-transporting ATP synthase (H
pir1:PWRTAM	+	63.50	116.40	33.66	509 H+-transporting ATP synthase (H
pir1:PWZAM	+	63.50	116.38	33.66	510 H+-transporting ATP synthase (H
pir2:T10957	+	63.50	116.38	33.66	510 H+-transporting ATP synthase (H
pir2:T19261	+	63.50	116.38	33.66	510 H+-transporting ATP synthase (H

pir2:B86244 + 63.50 112.12 33.52 884 ! unknown protein, 43418-38554
pir2:T14595 + 63.50 109.04 33.42 1317 ! polyprotein - maize retrotr
pir2:A86647 + 63.00 116.35 38.56 447 ! hypothetical protein yb1b [i

seq_name: pir2:T02205

seq_documentation_block:

Lu-ECAM-1 protein - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 01-Dec-2000
C:Accession: T02205; T02152; T02171
E:Elbte; R.C.; Widom, J.; Gruber, A.D.; Abdel-Chany, M.; Levine, R.; Goodwin, A.; Pau
submitted to the EMBL Data Library, April 1997
A:Description: Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial
A:Reference number: Z14590
A:Accession: T02205
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-905 <ELB>
A:Cross-references: EMBL:AF001261; NID:g2623762; PIDN:AAB86529.1; PID:g2623763
A:Experimental source: lung
A:Accession: T02152
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-793, 'SGSP', 'M', '862', 'RE', '865-866', 'Q', '868', 'AKVLELO', '876', 'OHO', '880', 'FQ', '<P>
A:Cross-references: EMBL:AF001263; NID:g2623766; PIDN:AAB86531.1; PID:g2623767
A:Experimental source: lung
A:Accession: T02171
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-792, 'ES', '<EL3>
A:Cross-references: EMBL:AF001262; NID:g2623764; PIDN:AAB86530.1; PID:g2623765
A:Experimental source: lung

alignment_scores:

Quality: 216.00 Length: 83
Ratio: 3.130 Gaps: 1
Percent Similarity: 83.133 Percent Identity: 54.217

alignment_block:

US-09-049-696-8 x T02205

Align seg 1/1 to: T02205 from: 1 to: 905

3 CAAGTGTGTCATCATCACACAGTCTTGGGGCCCTGTCAGCTCA 52
447 sGluLeuGluThrLeuSerAsnMetThrGlyTyArgPheAlaA 464
431 ArgSerGlyAlaIleIleHisThrIleAlaLeuGlyProSerAlaAla 447
53 AGAAGTGTGTCATCATCACAGGAGTTTACAGACATATGCTT 102
447 sGluLeuGluThrLeuSerAsnMetThrGlyTyArgPheAlaA 464
103 CAGATCAAGTTTCAACAACATCGCTTCATTGATGCTTTGGGGCCCTTCA 152
464 sLnYASpille.....ThrGlyLeuThrAsnAlaPheSerArgIleSer 478
153 TCAGGAATGAGCTGTCTCTCAGCGCTCCATCCAGCTTGGAGTAAGGG 202
479 SerArgSerGlySerIleThrGlnGlnAlaIleGlnLeuGluSerLysAl 495
203 ATTAACCTCCAGACACAGCCAGTGGATGAATGGCAGACATGATCGTGGAC 251
495 aLeuLysIleThrGlyArgLysArgValAsnGlyThrValProValasp 511

seq_name: pir1:A42489

seq_documentation_block:

glutamate dehydrogenase (NADP+) (EC 1.4.1.4) - Giardia lamblia
N:Alternate names: glutamic dehydrogenase; NADP-specific glutamate dehydrogenase
C:Species: Giardia lamblia
C:Date: 31-Dec-1993 #sequence_revision 23-Mar-1995 #text_change 11-Jun-1999
C:Accession: A42489

seq_name: p1r2:C81169

seq_documentation_block:

IgA-specific metalloendopeptidase (EC 3.4.24.13) NMB0700 [imported] - Neisseria mening

C:Species: Neisseria meningitidis

A:Variety: strain MC58 serogroup B; strain 81139

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C:Accession: C81169; S61326

R:Titellin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Tettelin, H.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. R.; Hickey, E.K.; Haft, D.H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M. ric, H.; Qin, H.;

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: AR1000; MUID:20175755

seq	name: pir2:C81169
-----	-------------------

```
seq documentation_block:
```

A: Variety: strain MC58 serogroup B; strain 81139
#text change 19-Jan-2001

C:\Date: "31-Mar-2000" #sequence=8, #files=1
C:\Accession: C81169; S61326

R. Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, B.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.

Hickey, E.K.; HALL, D.H.; SALZBERG, S.D.; WINTER, J.
HILL, H.: Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A;Authors: Grandt, G.; Sull, D.; Smith, H.O.; Fraser, J.
Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.

A; Reference number: A81000; MUID: 20175755

A;Accession: C81169

A: Residues: 1-1815 <TET>

A; Cross-references: GB: AE002098; NID: y/425225;
GB: AE002424; GB: AE002098 p strain MC58

A; Experimental source: Poulson, K.; Mogens, K. Poulson, K.; Mogens, K.

Mol. Microbiol. 15, 495-506, 1995

A; Title: Comparative Characterization of
A; Reference number: S61314; MUID: 95302961

A;Accession: S61326

A; Status: preliminary; nucleic acid sequence not available

A;Residues: 53-548 <LOM>

A;Cross-references: EMBL: X62477, NID: 9752050, 12886666

A:Gene: NMB0700

C; Superfamily: IgA-specific metalloendopeptidase

C:\keywofas: myuofuse, "ccatofpocent"

alignment_scores: 74.00
Length: 83

Ratio: 1.480

Percent Similarity: 60.241 Percent Identity: 31.522

alignment block:

US-09-049-696-8 x C81169 ..

Align seq 1/1 to: C81169 from: 1 to: 1815

.....CAGCCGACCTT 67

.....
18 ATCCACAGTCGCTTTGGGGCGGCAGCAGCAATGAT
.....
:::
:::

1496 ValGluThrValSerLeuGlnProArgAlaAlaGlnProArgAlaGlnAl 1512

60 CTCCCAAAATCAACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCAG. 116

Year	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	

15T2 aAlaThrGln.....LeuGLIAAGCATTATCATGGCTTTTC

117AACAAATGGCCTCATTGATGCTTTTGGGGCCCTT 149

Align seg 1/1 to: S74325 from: 1 to: 250

```
6 AGTGGTCCATCATCCACAGCTGCTTTGGGGCCCTCTGCAGCTCAAGA 55
|||||
27 SerGlyAlaIleLeuThrLeuProSerProLeuTrpAlaAsnG1 43
|||||
56 ACTAGAGGACCTGCCAAATGACAGGA..... 83
|||||
43 uMetGlnGlnLeuLeuThrValThrGlyGlnAlaSerGluSerValProt 60
|||||
84 .....GGTTTACAGACATATGCTTCA 104
|||||
60 hrThrLeuThrGluAlaGlnLeuThrValGluValGlnAlaAspThrAla 76
|||||
105 GATCAAGTTCCAGAACATAGCCTCATTTGCTTTGGGGCCCTTTCATC 154
|||||
77 AspGlnValGlnThr.....AlaIleAlaG1 85
|||||
155 AGGAATGAGGCTGTC.....TCTCAGCGCTCCATCCAGC 189
|||||
85 uArgSerAsnArgLeuValAlaTyrLeuArgSerGlnArgValAspLysL 102
|||||
190 TTGAGAGTAAGGATTAACTCCAGC 215
|||||
102 euGlnThrGlnGlyLeuGlnLeuGln 110
|||||
```

seq_name: pir2:S31831

```
seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) - Neisseria gonorrhoeae (strain NG74 an
N:Alternate names: Iga-specific proteinase
C:Species: Neisseria gonorrhoeae
A:Variety: strain NG74; strain R16
C:Date: 22-Nov-1993 #sequence_revision 23-Feb-1996 #text_change 08-Dec-2000
C:Accession: S31831; S31832; S40434; S40433
R:Klausner, T.
Submitted to the EMBL Data Library, February 1993
A:Reference number: S31831
A:Accession: S31831
A:Molecule type: DNA
A:Residues: 1-370 <KLA>
A:Cross-references: EMBL:221615; NID:g44866; PIDN:CAA79739.1; PID:g940787
A:Experimental source: strain NG74
A:Accession: S31832
A:Molecule type: DNA
A:Residues: 1-370 <KLM>
A:Cross-references: EMBL:221616; NID:g44867; PIDN:CAA79740.1; PID:g940788
A:Experimental source: strain R16
R:Klausner, T.; Kraemer, J.; Otzelberger, K.; Pohlner, J.; Meyer, T.F.
J. Mol. Biol. 234, 579-593, 1993
A:Title: Characterization of the Neisseria Iga(beta)-core. The essential unit for outer
A:Reference number: S40431; MUID:94076337
A:Accession: S40434
A:Molecule type: DNA
A:Status: nucleic acid sequence not shown
A:Residues: 1-5, 'L', 7-10, 'p', 12-20, 25-30, 'EE', 33-35, 'AARE', 36-370 <KLF>
A:Cross-references: EMBL:221616
A:Experimental source: strain R16
A:Accession: S40433
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-370 <KLG>
A:Cross-references: EMBL:221615; NID:g44866; PIDN:CAA79739.1; PID:g940787
A:Experimental source: strain NG74
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase
```

alignment_scores:
Quality: 69.00 Length: 77
Ratio: 1.438 Gaps: 2
Percent Similarity: 62.338 Percent Identity: 28.571

alignment_block:

```
US-09-049-696-8 x S31831 ..
Align seg 1/1 to: S31831 from: 1 to: 370
18 ATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGGAGCT 67
|||||
59 ValGluAlaValSerLeuGlnProArgAlaAlaGlnProArgThrGlnAl 75
|||||
68 GTCCAAATGACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTTCAGA 117
|||||
75 aAlaAlaGlnAlaAspAlaValSerThrAsnThr.....A 87
|||||
118 ACATGGCCCTCATTTGATGCTTTTGGGGCCCTTTCATCA.....GGA 158
|||||
87 snSerAlaLeuSerAspAlaMetAlaSerThrGlnSerIleLeuLeuAsp 103
|||||
159 AATGGAGCTGCTCTCAGCGCTCCATCCAGCTTCAGACTAAGGATTAAAC 208
|||||
104 ThrGlyAlaTyrLeuThrArgHisIleAlaGlnLysSerArgAlaAspAl 120
|||||
209 CTCCAGACAGCCAGTGGATGATGATGATGATGATGATGATGATGATGATG 239
|||||
120 aGluLysAsnSerValTrpMetSerAsnThr 130
|||||
```

seq_name: pir2:S40431

```
seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) - Neisseria meningitidis (strain B19
N:Alternate names: Iga-specific proteinase
C:Species: Neisseria meningitidis
A:Variety: strain B1939
C:Date: 07-May-1998 #sequence_revision 17-Jul-1998 #text_change 08-Dec-2000
C:Accession: S40431; S31833
R:Klausner, T.; Kraemer, J.; Otzelberger, K.; Pohlner, J.; Meyer, T.F.
J. Mol. Biol. 234, 579-593, 1993
A:Title: Characterization of the Neisseria Iga(beta)-core. The essential unit for out
A:Reference number: S40431; MUID:94076337
A:Accession: S40431
A:Molecule type: DNA
A:Status: nucleic acid sequence not shown
A:Residues: 1-370 <KLA>
A:Cross-references: EMBL:221613
A:Experimental source: strain B1939
R:Klausner, T.
Submitted to the EMBL Data Library, February 1993
A:Reference number: S31831
A:Accession: S31833
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5, 'SI', 8-10, 'A', 12-19, 'Y', 21-22, 'NDYE', 23-26, 'DA', 29-31, 36-249, 'R', 251-
A:Cross-references: EMBL:221613; NID:g45061; PIDN:CAA79737.1; PID:g940805
A:Experimental source: strain B1939
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase
```

alignment_scores:
Quality: 69.00 Length: 75
Ratio: 1.500 Gaps: 2
Percent Similarity: 61.333 Percent Identity: 29.333

alignment_block:

```
US-09-049-696-8 x S40431 ..
Align seg 1/1 to: S40431 from: 1 to: 370
18 ATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGGAGCT 67
|||||
59 ValGluThrValSerLeuGlnProArgAlaAlaGlnProArgAlaGlnAl 75
|||||
68 GTCCAAATGACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTTCAGA 117
|||||
```

```
75 alaalaGlnAlaAspAlaValSerThrAsnThr.....A 87
118 ACAATGGCGCTCATTTGCGGCCCTTTTCATCA.....CGA 158
87 snSerAlaLeuSerAspAlaMetAlaSerThrGlnSerIleLeuLeuAsp 103
159 AATGGAGCTGCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGGATTAAAC 208
104 ThrGlyAlaTyrLeuThrArgHisIleAlaGlnLysSerArgAlaAspAl 120
209 CCTCCAGACAGCCAGTGGATGATGCACA 239
1282 aGluLysAsnSerValTrpMetSerAsnThr 1292
seq_name: pir2:S73738
seq_documentation_block:
phosphoglycerate kinase (bc 2.7.2.3) p9k - Mycoplasma pneumoniae (strain ATCC 29342)
N:Alternate names: hypothetical protein A05_0r1409
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
R:Hilmeireich, R.; Hilbert, H.; Plagens, H.; Pirk, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885
A:Accession: S73738
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-409 <HIM>
A:Cross-references: EMBL:AE000040; GB:U00089; NID:gl674091; PIDN:AAB96060.1; PID:gl67
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Gene: p9k
A:Genetic code: SGC3
A:Superfamily: phosphoglycerate kinase
C:Keywords: ATP; phosphotransferase
alignment_scores:
Quality: 68.50 Length: 48
Ratio: 2.076 Gaps: 1
Percent Similarity: 68.750 Percent Identity: 35.417
alignment_block:
US-09-049-696-8 x S73738 ..
Align seg 1/1 to: S73738 from: 1 to: 409
24 ACAGTCGCTTTGGGCCCTTCACGCTCAAGACTAGAGAGCTGTCCTCA 73
346 ThrLeuAlaIleGlyAlaIleAlaGlnAsnGlnGlnAlaPheSerVa 362
74 AATGACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGACAAATG 123
362 lileGlyGlyGlyAspSerAlaAlaAlaLysGlnLeuGln..... 376
124 GCCTCATTCATGCTTTTGGGCCCTTCATCAGGAATGGAGCT 167
377 ..ileAlaAspGlnPheSerPheIleSerThrGlyGlyAla 390
seq_name: pir2:E82750
seq_documentation_block:
hemagglutinin-like secreted protein XF0889 [imported] - Xylella fastidiosa (strain 9a)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: E82750
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: E82750
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3282 <SIM>
```

```
75 alaalaGlnAlaAspAlaValSerThrAsnThr.....A 87
118 ACAATGGCGCTCATTTGCGGCCCTTTTCATCA.....CGA 158
87 snSerAlaLeuSerAspAlaMetAlaSerThrGlnSerIleLeuLeuAsp 103
159 AATGGAGCTGCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGGATTAAAC 208
104 ThrGlyAlaTyrLeuThrArgHisIleAlaGlnLysSerArgAlaAspAl 120
209 CCTCCAGACAGCCAGTGGATGATGCACA 233
120 aGluLysAsnSerValTrpMetSer 128
seq_name: pir2:A26039
seq_documentation_block:
Iga-specific metalloendopeptidase (EC 3.4.24.13) precursor - Neisseria gonorrhoeae (stra
N:Alternate names: Iga protease; immunoglobulin A1 proteinase
C:Species: Neisseria gonorrhoeae
A:Variety: strain MS11
C:Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 08-Dec-2000
C:Accession: A26039; S09386
R:Pohlner, J.; Halter, R.; Beyreuther, K.; Meyer, T.F.
Nature 325, 458-462, 1987
A:Title: Gene structure and extracellular secretion of Neisseria gonorrhoeae Iga protease
A:Reference number: A26039; MUID:87115823
A:Accession: A26039
A:Molecule type: DNA
A:Residues: 1-1532 <POH>
A:Cross-references: GB:X04835; NID:q44868; PIDN:CAA28538.1; PID:q44869
A:Note: the authors translated the codon AAG for residue 668 as Asn
R:Halter, R.; Pohlner, J.; Meyer, T.F.
EMBO J. 8, 2737-2744, 1989
A:Title: Mosaic-like organization of Iga protease genes in Neisseria gonorrhoeae generat
A:Reference number: S09386; MUID:90060036
A:Accession: S09386
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 281-325, 'N', 327-337, 'N', 339-427, 'M', 429-531, 'N', 533-615, 'V', 617-631, 'N', 633-
A:Experimental source: strain MS11
C:Genetics:
A:Gene: iga
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase; transmembrane protein
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-1532/Product: immunoglobulin A1 proteinase #status predicted <MAT>
F:986-987/Cleavage site: Pro-Ser (autolytic) #status predicted
F:1018-1019/Cleavage site: Pro-Ser (autolytic) #status predicted
F:1121-1122/Cleavage site: Pro-Ala (autolytic) #status predicted
alignment_scores:
Quality: 69.00 Length: 77
Ratio: 1.438 Gaps: 2
Percent Similarity: 62.338 Percent Identity: 28.571
alignment_block:
US-09-049-696-8 x A26039 ..
Align seg 1/1 to: A26039 from: 1 to: 1532
18 ATCCACACAGTCGTTTGGGCCCTTCACGCTCAAGACTAGAGGAGCT 67
1221 ValGluThrAlaGlyLeuGlnProArgAlaAlaGlnProArgThrGlnAl 1237
68 CTCCTCAAAATGACAGGAGTTTACAGACATATGCTTCAGATCAAGTTCAGA 117
1237 aAlaAlaGlnAlaAspAlaValSerThr.....A 1249
118 ACAATGGCGCTCATTTGCGGCCCTTTTCATCA.....CGA 158
1249 snSerAlaLeuSerAspAlaMetAlaSerThrGlnSerIleLeuLeuAsp 1265
```

A: Residues: 1-3442 <SIM>
A: Cross-references: GB:AE004032; GB:AE003849; NID:g9107324; PIDN:AAF84995.1; GSPDB:GN
A: Experimental source: strain 9a5c
R: Simpson, A.J.G.; Reinach, F.C.; Arruga, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
B: Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A: Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitzajima, J.P.; Krieger, J.E.; Kuranaka, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A: Authors: Martins, E.M.F.; Matsukuma, A.Y.; Mencia, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A: Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tshakoto, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovsky-Almeida, S.; Vettore, A.L.
A: Reference number: A59328
A: Contents: annotation
C: Genetics
A: Gene: XF2196

```

alignment_scores:
  Quality: 68.00
  Ratio: 1.259
  Percent Similarity: 48.214
  Length: 112
  Gaps: 3
  Percent Identity: 22.321

alignment_block:
  US-09_049-696-8 x E82589 ..
  Align seq 1/1 to: E82589 from: 1 to: 3442

```

```

Align seg I/I to: E92589 from: 1 to: 3442

      9 GTGGCCATCATCCACAGCTGCGTTGGGCCCTCGCAGGTCAAGAACT 58
        |||:::||:::||:::||| ||| ||||| ::|
768 GlyGlyIleLeuHisThr...AlaThrGlyProLeuAspLeuArgIleThr 783
      59 AGAGGAGCTGCTCCAAAATGACAGGAGGTTTTACAGACA..... 95
        |||:::||:::||:::||| ||| |||||
783 rGlyThrLeuAsnAsnGlnAspAsnGlyLeuLeuThrThrAspAlaL 800
      95 .....TATGCTTCAGATCAAGTTCAGACAATGGCTCATGTGCT 137
        |||:::||:::||:::||| ||| |||||
800 euThrLeuThrAlaAlaSerLeuLeuAsnGlnHisGlyThrLeuAspAla 816
      138 TTTGGG..... 143
          |||
817 AlaGlyProAlaHisLeuThrLeuThrGlyLeuLeuAspAsnThraLaGI 833
      144 .....GCCTTTTCAT 153
          |||:::||:::||:::|||
833 yLeuLeuGlnThrAlaHisThrLeuTrpLeuThrSerAlaGlyLeuThra 850
      154 CAGGAATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGGA 203
          ::| |||||:::||:::||| |||:::||:::|||
850 snArgSerGlyThrLeuThrAlaSerGlnLeuThrLeuAspThrGlnGly 866
      204 TTAA CCTCCAGAACGCCAGTGGATGAATGGCAC 239
          |||||:::||:::||:::||| |||||
867 HisThrLeuAspAsnThrArgGLyValrLeuGLyThr 878

```

```

seq_name: p1r2:E82589

seq_documentation_block:
hemagglutinin-like secreted protein XF2196 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: E82589
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: E82589
A>Status: preliminary
A:Molecule type: DNA

```

This Page Blank (uspio)

OM of: US-09-049-696-8 to: Issued_Patents_AA:* out_format : pfs

Date: Mar 30, 2002 2:33 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+2np.model -DEV=xlp
-O/cgn2_1/USPRO_spool/US09049696/runat_28032002_145238_2012/app_query.fasta_1.12579
-DB-Issued_Patents_AA -OPMT-fastan -SUFFIX=ra1 -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -OGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62
-TRANS=human40.cd1 -LIST=45 -DLOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1_109 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-8

Query length: 253

Database: Issued_Patents_AA:*

Database sequences: 212252

Database length: 22503292

Search time (sec): 362.110000

score_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-342-749-2		69.50	137.13	0.7066	492
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-342-749-2		68.00	122.73	1.46	1507
/cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-713-298B-2		66.00	130.97	1.92	400
/cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-870-180B-2		66.00	130.97	1.92	400
/cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-814-052A-4		66.00	130.97	1.92	400
/cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-812-829-4		66.00	130.97	1.92	400
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-226-529-2		66.00	130.97	1.92	400
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-870-180B-13		66.00	129.57	1.98	462
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-226-529-13		66.00	129.57	1.98	462
/cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-09-310-293-2		64.50	116.54	3.96	1229
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-188-930-125		61.50	122.33	7.03	330
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-188-930-303		61.50	116.24	8.21	617
/cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-818-112-114		60.00	117.77	11.32	368
/cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-09-055-097-1		59.50	117.50	11.39	378
/cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-727-548-2		59.50	114.28	13.95	467
/cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-09-185-160-11		59.50	109.92	15.58	731
/cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-477-396A-4		59.00	108.16	18.38	777
/cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-333-565-59		58.50	118.71	15.84	233
/cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-081-448-6		58.50	118.71	15.84	233
/cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-607-269-24		58.50	118.71	15.84	233
/cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-471-058-14		58.50	118.71	15.84	233
/cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-661-479-59		58.50	118.71	15.84	233
/cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-470-670A-7		58.50	118.71	15.84	233
/cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-471-057-14		58.50	118.71	15.84	233
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-481-739-2		58.50	118.71	15.84	233
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-167-921-2		58.50	118.71	15.84	233
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-323-743-2		58.50	118.71	15.84	233
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-461-511A-7		58.50	118.71	15.84	233
/cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US94-07089-7		58.50	118.71	15.84	233
/cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US95-04600-24		58.50	118.71	15.84	233
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-751-359-22		58.00	117.71	18.32	229
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-123-492A-1		57.50	118.41	20.29	189
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-254-733-7		57.50	116.80	21.14	223
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-101-519-1		57.50	116.71	21.19	225
/cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-922-635-22		57.50	101.54	31.17	1070
/cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-798-897-18		56.50	124.08	22.33	83
/cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-978-523-18		56.50	124.08	22.33	83
/cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-121-436A-2		56.50	114.46	28.52	223
/cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-07-809-457A-16		56.50	110.42	31.62	338
/cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-553-943-16		56.50	110.42	31.62	338

/cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US91-09437-16 + 56.50 110.42 31.62 453
/cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-07-937-609-26 - 56.50 107.57 33.99 453
/cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-029-170-26 - 56.50 107.57 33.99 453
/cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-451-715A-8 - 56.50 101.20 39.98 872
/cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-09-045-632-29 + 56.00 115.27 31.50 182

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-342-749-2

seq_documentation_block:

Sequence 2, Application US/09342749
Patent No. 6166194
GENERAL INFORMATION:
APPLICANT: Wong, Alexander K.C.
APPLICANT: Tavtigian, Sean V.
APPLICANT: Teng, David H.-F.
APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: TMPS2 is a Tumor Suppressor
FILE REFERENCE: 2318-202
CURRENT APPLICATION NUMBER: US/09/342,749
CURRENT FILING DATE: 1999-06-29
EARLIER APPLICATION NUMBER: US 60/091,044
EARLIER FILING DATE: 1998-06-29
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 492
TYPE: PRT
ORGANISM: Homo sapiens
US-09-342-749-2

alignment_scores:

	Quality:	69.50	Length:	77
	Ratio:	1.580	Gaps:	3
Percent Similarity:	57.143	Percent Identity:	29.870	
alignment_block:				
US-09-049-696-8 x US-09-342-749-2				
Align seg 1/1 to: US-09-342-749-2 from: 1 to: 492				
11	GGCCATCATCCACACAGTCGCTTTGGGCGCTCTGCAGCTCAAGACTAG	60		
111				
65	CysThrGlnProLysSerProSerGlyThrValCysThrSerLysThrLys	82		
61	AGGAGCTGTCCAAATACAGAGAGGTTTACAGACATATGCTTCAGATCAA	110		
82	sLysAlaLeu.CysIleThrLeuThrLeuGlyThrPheLeuValGlyAla	98		
111	GTTGAGACAAATGCCCTCATTCATGCTTTT...GGGCGCTTTTCATCAGG	157		
93	AlaLeuAlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAs	115		
156	AAATGAGAGCTGCTCTCTCAGCGCTCCATCCAGCTTGAGAGTAAGGATTAA	207		
115	nSerGly.....IleGluCysAspSerGlyThrC	126		
209	CCCTCCAGAACAGCCAGTGGATGAATGCC	236		
125	ysIleAsnProSerAsnTrpCysAspGly	135		

seq_name: /cgn2_6/ptodata/2/1aa/backfiles1.pep:5268270-2

seq_documentation_block:

Patent No. 5268270
APPLICANT: Meyer, Thomas F.; Halter, Roman; Pohlner, Johannes
TITLE OF INVENTION: PROCESS FOR PRODUCING PROTEINS USING GRAM
NEGATIVE HOST CELLS
NUMBER OF SEQUENCES: 6
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/171,872
FILING DATE: 01-JUL-1987
SEQ ID NO: 2

```

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 400 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-713-298B-2
;
alignment_scores:
  Quality: 66.00      Length: 46
  Ratio: 2.062       Gaps: 2
  Percent Similarity: 69.565      Percent Identity: 34.783
;
alignment_block:
  US-09-049-696-8 x US-08-713-298B-2 ..
  Align seg 1/1 to: US-08-713-298B-2 from: 1 to: 400
;
  93 ACATATGCTTCAGATCAAGTTCAGACAATGGCGCTCATGTCATGCTTTGG 142
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  23 ThrThrAlaAlaAsp.....AsnAspSerValValGluGluHisG1 36
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  143 GCGCCCTTCATCAGAAATGGA...GCTGCTCTCTCAGCGCTCCATCCAGC 189
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  36 yclnLeuSerIleSerAsnGlyGluLeuValAsnGluArgGlyGluGlnv 53
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  190 TTGAGAGTAAGGATTAACTCCAGACAGCCAGTGG 227
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  53 aiglnLeuLysGlyMetSerSerHisGlyLeuGlnTrp 65
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
;
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-713-298B-2
;
seq_documentation_block:
; Sequence 2, Application US/08870180B
; Patent No. 5945327
; GENERAL INFORMATION:
; APPLICANT: Outtrup, Helle
; APPLICANT: Dambmann, Claus
; APPLICANT: Olsen, Arne
; APPLICANT: Bisg rd-Frantzen, Henrik
; APPLICANT: Sch lein, Martin
; APPLICANT: J rgensen, Per
; APPLICANT: Bjoernvad, Mads
; TITLE OF INVENTION: DNA Constructs and Methods of Producing
; TITLE OF INVENTION: Cellulytic Enzymes
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 5945327o No. 5945327disk of No. 5945327th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,180B
; FILING DATE: 6-June-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol E.
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 3794.434-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 400 amino acids
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1507
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-713-298B-2
;
alignment_scores:
  Quality: 68.00      Length: 77
  Ratio: 1.417       Gaps: 2
  Percent Similarity: 62.338      Percent Identity: 28.571
;
alignment_block:
  US-09-049-696-8 x 5268270-2 ..
  Align seg 1/1 to: 5268270-2 from: 1 to: 1507
;
  18 ATCCACACAGTCGCTTTGGGCGCTTCGACGCTCAAGACTAGAGGAGCT 67
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  1196 ValGluThrAlaGlyLeuGlnProArgAlaAlaGlnProArgThrGlnAl 1212
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  68 GTCCAAAATGACAGAGGTTTACAGACATGCTTCAGATCAAGTTCAGA 117
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  1212 aaAlaAlaGlnAlaAspAlaValSerThrAsnThr.....A 1224
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  118 ACAATGGCGCTCATGCTTTTGGGCGCTTTCATCA.....GGA 158
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  1224 snSerAlaLeuSerAspAlaMetAlaSerThrGlnSerIleLeuLeuAsp 1240
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  159 AATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGAGTAGTAGGATTAC 208
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  1241 ThrGlyAlaTyLeuThrArgHisIleAlaGlnLysSerArgAlaAspAl 1257
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  209 CTTCCAGAACACCCAGTGGATGAATGGCACA 239
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  1257 aGluLysAsnSerValTrpMetGluAsnThr 1267
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
;
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-713-298B-2
;
seq_documentation_block:
; Sequence 2, Application US/087113298B
; Patent No. 5922586
; GENERAL INFORMATION:
; APPLICANT: Outtrup, Helle
; APPLICANT: Dambmann, Claus
; APPLICANT: Olsen, Arne
; APPLICANT: Bisg rd-Frantzen, Henrik
; APPLICANT: Sch lein, Martin
; APPLICANT: J rgensen, Per
; APPLICANT: Bjoernvad, Mads
; TITLE OF INVENTION: DNA Constructs and Methods of Producing
; TITLE OF INVENTION: Cellulytic Enzymes
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 5922586o No. 5922586disk of No. 5922586th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/713,298B
; FILING DATE: 13-SEPT-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valeta A.
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 3794.424-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 867-0123

```


alignment_scores:		
Quality:	66.00	Length: 46
Ratio:	2.062	Gaps: 2
Percent Similarity:	69.565	Percent Identity: 34.783

Tue Apr 2 09:40:14 2002

us-09-049-696-8.rai

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-188-930-125

seq_documentation_block:

; Sequence 125, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188.930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 125
; LENGTH: 330
; TYPE: PRT
; ORGANISM: mouse
US-09-188-930-125

alignment_scores:
Quality: 61.50 Length: 55
Ratio: 1.864 Gaps: 3
Percent Similarity: 60.000 Percent Identity: 38.182

alignment_block:

US-09-049-696-8 x US-09-188-930-125 ..
Align seg 1/1 to: US-09-188-930-125 from: 1 to: 330

34 TGGGGCCCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAAATGACAGGA 83
||| ||||| : : : : : ||||| : : : : :
55 TrpPheProLeuLeuValArgLysSerProLysPheGluAs 71
84 GGTTCACACATATGCTTCAGATCAAGTTTCAGAACAAATGGCCTCATTTGA 133
||| ||||| : : : : : ||||| : : : : :
71 pTyTrrpArgHis.....ArgThrProAlaSerPheG 82
134 TG.....CTTTTGGGG...CCCTTTTCATCAGAAATGGAGCTGTCTCTC 174
||| ||||| : : : : : ||||| : : : : :
82 InArgLysLeuLeuGlySerProSerLeuSerGluSerHisArgIle 98
175 AGCGCTCCATCCAGC 189
||| ||||| : : : : :
99 SerIleProSerSer 103

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-188-930-303

seq_documentation_block:

; Sequence 303, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188.930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 303
; LENGTH: 617
; TYPE: PRT

ORGANISM: Mouse
US-09-188-930-303

alignment_scores:

Quality: 61.50 Length: 55
Ratio: 1.864 Gaps: 3
Percent Similarity: 60.000 Percent Identity: 38.182

alignment_block:

US-09-049-696-8 x US-09-188-930-303 ..
Align seg 1/1 to: US-09-188-930-303 from: 1 to: 617
34 TGGGGCCCTCTGCAGCTCAAGAACTAGAGGAGCTGTCCAAAATGACAGGA 83
||| ||||| : : : : : ||||| : : : : :
55 TrpPheProLeuLeuValArgLysSerProLysPheGluAs 71
84 GGTTCACACATATGCTTCAGATCAAGTTTCAGAACAAATGGCCTCATTTGA 133
||| ||||| : : : : : ||||| : : : : :
71 pTyTrrpArgHis.....ArgThrProAlaSerPheG 82
134 TG.....CTTTTGGGG...CCCTTTTCATCAGAAATGGAGCTGTCTCTC 174
||| ||||| : : : : : ||||| : : : : :
82 InArgLysLeuLeuGlySerProSerLeuSerGluSerHisArgIle 98
175 AGCGCTCCATCCAGC 189
||| ||||| : : : : :
99 SerIleProSerSer 103

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-818-112-114

seq_documentation_block:

; Sequence 114, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid

This Page Blank (uspto)

seq_documentation_block:

us-09-049-696-8.rag

Tue Apr 2 09:40:13 2002

```
PR 24-JUN-1998; 98US-0090429;
PR 24-JUN-1998; 98US-0090431;
PR 24-JUN-1998; 98US-0090435;
PR 24-JUN-1998; 98US-0090445;
PR 24-JUN-1998; 98US-0090451;
PR 24-JUN-1998; 98US-0090461;
PR 24-JUN-1998; 98US-0090535;
PR 24-JUN-1998; 98US-0090540;
PR 24-JUN-1998; 98US-0090557;
PR 25-JUN-1998; 98US-0090676;
PR 25-JUN-1998; 98US-0090678;
PR 25-JUN-1998; 98US-0090688;
PR 25-JUN-1998; 98US-0090690;
PR 25-JUN-1998; 98US-0090691;
PR 25-JUN-1998; 98US-0090694;
PR 25-JUN-1998; 98US-0090695;
PR 25-JUN-1998; 98US-0090696;
PR 26-JUN-1998; 98US-0090862;
PR 26-JUN-1998; 98US-0090863;
PR 01-JUL-1998; 98US-0091358;
PR 01-JUL-1998; 98US-0091360;
PR 01-JUL-1998; 98US-0091544;
PR 02-JUL-1998; 98US-0091478;
PR 02-JUL-1998; 98US-0091486;
PR 02-JUL-1998; 98US-0091519;
PR 02-JUL-1998; 98US-0091626;
PR 02-JUL-1998; 98US-0091628;
PR 02-JUL-1998; 98US-0091633;
PR 02-JUL-1998; 98US-0091646;
PR 02-JUL-1998; 98US-0091673;
PR 02-JUL-1998; 98US-0091978;
PR 07-JUL-1998; 98US-0091982;
PR 09-JUL-1998; 98US-0092182;
PR 10-JUL-1998; 98US-0092472;
PR 20-JUL-1998; 98US-0093339;
PR 30-JUL-1998; 98US-0094651;
PR 04-AUG-1998; 98US-0095285;
PR 04-AUG-1998; 98US-0095301;
PR 04-AUG-1998; 98US-0095302;
PR 04-AUG-1998; 98US-0095318;
PR 04-AUG-1998; 98US-0095321;
PR 04-AUG-1998; 98US-0095325;
PR 10-AUG-1998; 98US-0095916;
PR 10-AUG-1998; 98US-0095929;
PR 10-AUG-1998; 98US-0096012;
PR 11-AUG-1998; 98US-0096143;
PR 11-AUG-1998; 98US-0096146;
PR 12-AUG-1998; 98US-0096329;
PR 17-AUG-1998; 98US-0096757;
PR 17-AUG-1998; 98US-0096766;
PR 17-AUG-1998; 98US-0096768;
PR 17-AUG-1998; 98US-0096773;
PR 17-AUG-1998; 98US-0096773;
PR 17-AUG-1998; 98US-0096791;
PR 17-AUG-1998; 98US-0096867;
PR 17-AUG-1998; 98US-0096891;
PR 17-AUG-1998; 98US-0096894;
PR 17-AUG-1998; 98US-0096895;
PR 17-AUG-1998; 98US-0096897;
PR 18-AUG-1998; 98US-0096949;
PR 18-AUG-1998; 98US-0096950;
PR 18-AUG-1998; 98US-0096950;
PR 18-AUG-1998; 98US-0096959;
PR 18-AUG-1998; 98US-0097022;
PR 19-AUG-1998; 98US-0097141;
PR 20-AUG-1998; 98US-0097218;
PR 20-AUG-1998; 98US-0097661;
PR 26-AUG-1998; 98US-0097951;
PR 26-AUG-1998; 98US-0097952;
PR 26-AUG-1998; 98US-0097954;
PR 26-AUG-1998; 98US-0097955;

PR 26-AUG-1998; 98US-0097971;
PR 26-AUG-1998; 98US-0097974;
PR 26-AUG-1998; 98US-0097978;
PR 26-AUG-1998; 98US-0097979;
PR 26-AUG-1998; 98US-0097986;
PR 26-AUG-1998; 98US-0098014;
PR 31-AUG-1998; 98US-0098525;
PR 16-SEP-1998; 98US-0100634;
PR 12-JAN-1999; 99US-0115565;
XX (GETH ) GENENTECH INC.
XX Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
XX Wood WI, Yuan J;
XX WPI: 2000-072883/06.
XX N-PSDB: AAZ65095.
XX Membrane-bound proteins and related nucleotide sequences
XX claim 12: Fig 274; 822pp; English.
XX The invention provides membrane-bound PRO polypeptides and
XX polynucleotides encoding them. The PRO sequences of the invention were
XX identified based on extracellular domain homology screening. The PRO
XX sequences have homology with proteins including LDL receptors, TIE
XX ligands and various enzymes. The membrane-bound proteins and receptor
XX molecules are useful as pharmaceutical and diagnostic agents. Receptor
XX immunoadhesins, for instance, can be used as therapeutic agents to block
XX receptor-ligand interactions. The membrane-bound proteins can also be
XX employed for screening of potential peptide or small molecule inhibitors
XX of the relevant receptor/ligand interaction. The PRO encoding sequences
XX are useful as hybridization probes, in chromosome and gene mapping and in
XX the generation of antisense RNA and DNA. PRO nucleic acid sequences
XX will also be useful for the preparation of PRO polypeptides, especially
XX by recombinant techniques.
XX Sequence 919 AA;
SQ
alignment_scores:
Quality: 276.00 Length: 83
Ratio: 3.833 Gaps: 0
Percent Similarity: 86.747 Percent Identity: 63.855
alignment_block:
US-09:049-696-8 x AAY66749
Align seg 1/1 to: AAY66749 from: 1 to: 919
3 CAAAGTGGTCCATCATCCACACAGTCGCTTTGGGGCCCTCTGCAGCTCA 52
|||||
428 GlnSerGlyAlaIleValHisPheIleAlaLeuGlyArgAlaAlaAspG1 444
53 AGAACTAGAGGAGCTCCAAATGACAGGAGGTTTACAGACATATGCTT 102
|||||
444 uAlaValLeuGluMetSerLysIleThrGlyGlySerHisPheThrValS 461
103 CAGATCAAGTTTCAGAACATGCCCTCATGATGCTTTGGGGCCCTTCA 152
|||||
461 erAspGluAlaGlnAsnAsnGlyLeuIleAspAlaPheGlyAlaLeuThr 477
153 TCAGGAAATGGAGCTGCTCTCAGGCTCCATCCAGCTTGAGAGTAAGG 202
|||||
478 SerGlyAsnThrAspLeuSerGlnLysSerLeuGlnLeuGluSerLysG1 494
203 ATTAACTCCAGACACGCCAGTCGATGAATGCACAGTCATCGTGGAC 251
|||||
494 yLeuThrLeuAsnSerAsnAlaTrpMetAsnAspThrValIleIleAsp 510
seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB87560
seq_documentation_block:
```


seq_name: /SIDS2/gcgdata/geneseq/AA2001.DAT:AAAG73761

seq_documentation_block:
ID AAG73761 standard; Protein; 218 AA.

AC AAG73761;

XX 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO:4525.

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.

OS Homo sapiens.

PN WO200122920-A2.

PD 05-APR-2001.

PF 28-SEP-2000; 2000WO-US26524.

PR 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Barash SC, Birse CE, Rosen CA;

DR WPI; 2001-235357/24.

DR N-PSDB; AAH33192.

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -

PS Claim 11; Page 6321; 9803pp; English.

CC AAH32943 to AAH37195 and AG73514 to AG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC present invention.

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 218 AA;

alignment_scores:
Quality: 265.00 Length: 83
Ratio: 3.786 Gaps: 0
Percent Similarity: 84.337 Percent Identity: 61.446

alignment_block:

US-09-049-696-8 x AAG73761

Align seg 1/1 to: AAG73761 from: 1 to: 218

3 CAAAGTGGTCCATCATCCACAGACAGTGGCTTTGGGGCCCTCTCAGCTCA 52
|||||

48 GlnSerGlyAlaIleValHisPheIleAlaLeuGlyArgAlaAlaAspG1 64
53 AGAACTAGAGGAGCTGCTCCAAATGACAGAGGTTTACAGACATATGCTT 102
64 uAlaValIleGluMetSerLysIleThrGlyGlySerHisPheTyrValS 81
105 CAGATCAAGTTTCAGAAACAATGGCTCATTCATGCTTTTGGGGCCCTTCA 152
81 erAspGluAlaGlnAsnAsnGlyLeuIleAspAlaPheGlyAla**Thr 97
153 TCAGGAAATGGAGTGTCTCTCAGCGCTCCATCCAGCTTCAGAGTAAGGG 202
99 SerGlyAsnThrAspLeuSer**LysSerIleuGlnLeuGluSerLysG1 114
203 ATTAACCTCCAGAACAGCCAGTGGATGAATGGCACAGTATCGTGGAC 251
114 yLeuThrLeuAsnSerAsnAlaTrpMetAsnAspThrValIleIleAsp 130
seq_name: /SIDS2/gcgdata/geneseq/AA1999.DAT:AAAY41039

seq_documentation_block:

ID AAAY41039 standard; protein; 592 AA.

AC AAAY41039;

DT 07-DEC-1999 (first entry)

XX Human lung tumor antigen L762P variant 1.

XX Human; lung tumor; lung cancer; T cell stimulation.

OS Homo sapiens.

PN WO9947674-A2.

PD 23-SEP-1999.

PE 17-MAR-1999; 99WO-US05798.

PR 18-MAR-1998; 98US-0040802.

PR 18-MAR-1998; 98US-0040984.

PR 27-JUL-1998; 98US-0123912.

XX 27-JUL-1998; 98US-0123933.

FA (CORI-) CORIXA CORP.

XX Reed SG, Wang T;

DR WPI; 1999-571839/48.

DR N-PSDB; AAZ24657.

XX New isolated lung tumor polynucleotides, used to develop products for
XX the treatment, prevention and monitoring the progression of lung cancer

PS Example 3; Page 144-145; 148pp; English.

CC The invention provides isolated human lung tumor nucleic acids and
CC polypeptides. The polypeptides can be used for the treatment of lung
CC cancer. The polypeptides and polynucleotides can be used to stimulate T
CC cells or antigen presenting cells for use in the treatment of lung
CC cancer. The polypeptides and monoclonal antibodies specific for the
CC polypeptides can also be used to inhibit the development of lung cancer.
CC Agents which bind the polypeptides can be used for detecting lung cancer
CC and for monitoring the progression of lung cancer.

SQ Sequence 592 AA;

alignment_scores:
Quality: 194.00 Length: 82
Ratio: 2.985 Gaps: 0
Percent Similarity: 79.268 Percent Identity: 48.780

```
alignment_block:
US-09-049-696-8 x AAY41039  ..
```

6 ATGTGTCATCATCCACACAGCTGGTTCGGGCCCTCTGCAGCTCAAGA 55
||||| :|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
436 SerGlySerThrIleHisSerIleAlaLeuGlySerSerAlaAlaProAs 452
||||| :|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::
56 ACTAAGAGGAGTGTCCTCAAATGACAGGAGGTTTACAGCATATGCTTCAG 105
||||| :|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::

106 ATCAAGTTCAGAACAATGGCTCATTCATGCTTGTGGGCCCTTTCATCA 155
|| : ::::|::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|:
469 spIleSerAsnSerAsnMetIleAspLaPheSerArgIleSerSer 485

156 GGAATGGAGCTGTCTCTCAGCGGCTCCATCCAGCTTGAGAGTAAGGGATT 205
|||||::|||
189 atctcgttcaatccacgagcgagggcgttgaggcttcgcctgctggaacgctggtg 502
|||||::|||

206 AACCTCCAGAACAGCCAGTGGATGAATGGCACAGTGATCGTGGAC 251

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA11324

seq_documentation_2008.
ID AAB11324 standard; Protein; 592 AA.
XX
XX

XX 21-FEB-2001 (first entry)
DT
XX

XX Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
KW screening; detection

XX
OS Homo sapiens.
XX

XX
PD
XX
19-OCT-2000.

XX
PR. 02-APR-1993;
99US-0285479.
00RG-0466306

PR 30-DEC-1999; 99US-0476490.
PR 10-JAN-2000; 2000US-0480884.
PR 22-FEB-2000; 2000US-0510376.

PA (CORI-) CORIXA CORP.
XX
PT Wang T. Fan L:

AA WPI; 2000-628399/60.
 DR N-PSDB; AAC65896.
 DR

PT Isolated polypeptide comprising an immunogenic portion of a protein of lung cancer

PT protein is used for detecting and monitoring progression of lung cancer

PT in a patient -

Claim 3; Page 177-178; 261pp; English.

comprising an immunogenic portion of a virus or virus-like particle (VLP) comprising one or more polypeptides and polynucleotides are which have cytostatic activity. The polypeptides and polynucleotides are used in compositions and vaccines to inhibit the development of cancer, Methods described in the invention

CC can be used to monitor the progression of a cancer by carrying out the

[illegible]

seq	name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT-AA70459
used	documentation_block:
ID	AA70459 standard; Protein; 942 AA.
XX	..
XX	AA70459;
XX	
XX	21-JUN-2000 (first entry)
DT	..
XX	Human membrane channel protein-9 (MECHP-9).
DE	..
XX	Membrane channel protein-9; MECHP-9; diagnosis; treatment: lymphoma;
XX	cell proliferative disorder; bursitis; atherosclerosis; cancer; sarcoma;
KW	inflammatory disorder; AIDS; Addison's disease; cystic fibrosis; asthma;
KW	diabetes mellitus; osmoregulatory disorder; diarrhoea; renal failure;
KW	muscular disorder; myocarditis; Duchenne's muscular dystrophy; nontropic;
KW	cardiovascular disorder; hypertension; bronchitis; vasculitis; cardiac;
KW	neurological disorder; Alzheimer's disease; Parkinson's disease; human;
KW	Funtington's disease; antiatherosclerotic; hepatotropic; cytostatic;
KW	anti-HIV; antandemic; neuroprotective; immunomodulator; antidiabetic;
KW	hypotensive; vasotropic; antistimatic; antiinflammatory; antidepressant;
KW	antiparkinsonian; immunostimulant.

XX	OS	Homio sapiens.	Location/Qualifiers
XX	PH	Key	
FT	FT	Modified-site	273 /note= "Phosphorylation site"
FT	FT	Modified-site	302 /note= "Phosphorylation site"
FT	FT	Modified-site	355 /note= "Phosphorylation site"
FT	FT	Modified-site	368 /note= "Phosphorylation site"
FT	FT	Modified-site	418 /note= "Phosphorylation site"
FT	FT	Modified-site	419 /note= "Phosphorylation site"
FT	FT	Modified-site	474 /note= "Phosphorylation site"
FT	FT	Modified-site	498 /note= "Phosphorylation site"
FT	FT	Modified-site	797 /note= "Phosphorylation site"
FT	FT	Modified-site	2 /note= "Phosphorylation site"
FT	FT	Modified-site	76 /note= "Phosphorylation site"
FT	FT	Modified-site	92 /note= "Phosphorylation site"
FT	FT	Modified-site	270 /note= "Phosphorylation site"
FT	FT	Modified-site	318 /note= "Phosphorylation site"
FT	FT	Modified-site	384 /note= "Phosphorylation site"
FT	FT	Modified-site	557 /note= "Phosphorylation site"
FT	FT	Modified-site	568 /note= "Phosphorylation site"
FT	FT	Modified-site	593 /note= "Phosphorylation site"
FT	FT	Modified-site	652 /note= "Phosphorylation site"
FT	FT	Modified-site	695 /note= "Phosphorylation site"
FT	FT	Modified-site	822 /note= "Phosphorylation site"
FT	FT	Modified-site	823 /note= "Phosphorylation site"
FT	FT	Modified-site	858 /note= "Phosphorylation site"
FT	FT	Modified-site	2 /note= "Phosphorylation site"

```

XX SO Sequence          942 AA:
aligned_t_scores:
    Quality: 194.00      Length: 82
    Ratio:   2.985       Gaps: 0
    Percent Similarity: 79.268     Percent Identity: 48.780

alignment_block:
US-09-049-696-8 x AAY70459 ..

Align seg 1/1 to: AAY70459 from: 1 to: 942

..6 AGTGGTGCCATCATCCACAGCAGTCGCTTTGGGGCCCTCTGCAGCTCAAGA 55
435 SerGlySerThrIleHisSerIleAlaLeuGlySerSerAlaIalaProAs 451
..56 ACTAGGAGGACTGTCCAAATGATGACAGSAGGTTTACAGACATATGCTTCAG 105
451 nLeuGluLeuSerArgLeuThrGlyGlyLeuLysPhePheValProA 468
106 ATCAAGATTACAGACAATGGCTCATTCATGATCGCTTTGGGGCCCTTCATCA 155
468 spIleSerAsnSerAsnSerMetIleAspAlaPheSerArgIleSerSer 484
156 GGAAATGAGACTGCTCTCTCACGCCCTCCATCCAGCTTCAGAGTAAGGATT 205
485 GlyThrGlyAspIlePheGlnHisIleGlnLeuGluSerThrGlyGI 501
206 AACCTCTCCAGAACAGCCAGTGGATGAATGGCAGAGTGATCGTGGAC 251
501 uAsnValLysProHisHisGlnLeuLysAsnThrValThrValasp 516

seq_name: /SDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AA41036

seg_documentation_block:
ID AA41036 standard; protein; 943 AA.
AC AAY41036;
XX
XX 07-DEC-1999 (first entry)
XX Human lung tumor antigen L762P.
XX Human; lung tumor; lung cancer; T cell stimulation.
XX Homo sapiens.
XX WO9947674-A2.
XX
XX PD 23-SEP-1999.
XX PF 17-MAR-1999; 99WO-US05798.
XX PR 18-MAR-1998; 98US-0040802.
XX PR 18-MAR-1998; 98US-0040984.
XX PR 27-JUL-1998; 98US-0123912.
XX PR 27-JUL-1998; 98US-0123933.
XX
XX (CORI-) CORIXA CORP.
XX PA Reed SG, Wang T;
XX PT WPT; 1999-571839/48.
XX DR N-PSDB; AA224653.
XX
XX PT New isolated lung tumor polynucleotides, used to develop o
XX PT the treatment, prevention and monitoring the progression o
XX
XX Example 3; Page 137-139; 148pp: English.

```


XX	WO9944620-A1.	
PN		
XX	10-SEP-1999.	
XX		
PD		
XX		
PF	03-MAR-1999; 99WO-US04703.	
XX		
PF	03-MAR-1998; 98US-0076815.	
PR		
XX		
PA	(MAGA-) MAGAININ PHARM INC.	
XX		
XX	Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;	
PI	Nicolaïdes NC, Zhou Y, Dong Q;	
PI		
XX	WPI: 1999-550979/46.	
DR	N-PSDB; AAF81926.	
DR		
XX		
XX	New nucleic acid encoding calcium activated chloride channel, used to	
PT	identify, e.g. specific modulators for treating atopic allergy -	
PT		
XX	Claim 13; Fig 4A; 75pp; English.	
XX		
XX	The present sequence represents the human interleukin 9 (IL-9) induced	
CC	calcium activated chloride channel 2 (ICACC-2) protein. ICACC proteins	
CC	have anti-allergic, anti-asthmatic, anti-inflammatory and	
CC	immunomodulatory activities. Compounds (A) that downregulate ICACC are	
CC	used to alleviate asthma (or more generally atopic allergy), while those	
CC	(B) that upregulate ICACC are used to treat cystic fibrosis (CF) and	
CC	inflammatory bowel disease (IBD) (or other autoimmune diseases).	
CC	Measuring levels of ICACC-1 can be used in the diagnosis of asthma	
CC	(increased levels) or IBD (reduced levels), also for monitoring	
CC	treatment of these conditions. The ICACC proteins can be used:	
CC	(i) to raise specific antibodies (Ab), useful: (a) as immunoassay	
CC	reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as	
CC	(A) to disrupt bonding between ICACC and its ligands; and (iii) to	
CC	identify modulators and binding partners. ICACC polynucleotides can be	
CC	used to generate transgenic animals or recombinant cells, used to screen	
CC	for antagonists, also as a source of therapeutic antisense agents or	
CC	diagnostic probes (for quantifying mRNA expression, e.g. for	
CC	identification of modulators).	
XX		
SQ	Sequence 943 AA;	

```

alignment_scores:
  Quality: 189.00      Length: 82
  Ratio: 2.908         Gaps: 0
  Percent Similarity: 79.268  Percent Identity: 47.561

alignment_block:
  US-09-049-696-8 x AAB74823      ..

  Align seg 1/1 to: AAB74823 from: 1 to: 943

6  ACTGGTGCCATCATCCACACAGCTGCTTGGGGCCCTCGCAGCTCAAGA 55
   |||||::: |||||:::|||||:::|||||::: |||||::: |||||:::
436 SerGlySerThrIleHisSerIleAlaLeuGlySerSerAlaAlaProAs 452
   |||||::: |||||:::|||||:::|||||::: |||||::: |||||:::
56  ACTAGAGGAGCTCTCCAAATACAGAGGCTTACAGACATATGCTTCAG 105
   :|||:::|||||:::|||||:::|||||::: |||||::: |||||:::
452 nLeuHisGluLeuSerArgLeuThrGlyGlyLeuLysPheValProA 469
   |||||::: |||||:::|||||:::|||||::: |||||::: |||||:::
106 ATCAAGTTCAGAACATATGGCTCATTCATGCTTTTGGGGCCCTTCATCA 155
   || |||||:::|||||:::|||||:::|||||::: |||||::: |||||:::
469 spileSerAsnSerAsnSerMetIleAspAlaPheSerArgIleSerSer 485
   |||||::: |||||:::|||||:::|||||::: |||||::: |||||:::
156 GGAATATGGAGCTCTCTCTCAGCGCTCCATCCAGCTTCAGAGCTAAGGGAT 205
   |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
486 GlyThrGlyAspIlePheGlnHisIleGlnLeuGluSerThrGlyGly 502
   |||||::: |||||:::|||||:::|||||::: |||||::: |||||:::
206 AACCCCTCCAGACAGCCAGCTGGATGAATGGCACAGCTGATCGTGGAC 251
   |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
502 uAsnValIysProHisIleGlnLeuLysAsnThrValThrValAsp 517
   |||||::: |||||:::|||||:::|||||::: |||||::: |||||:::

```

seq_name: ^b/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA01295

seq_documentation_block:

ID AAM01295 standard; Protein; 209 AA.

AA:01295; AC

04-OCT-2001 (first entry)

Human transmembrane protease serine 2 partial amino acid sequence.

Human: prostate cancer: prostate-specific: diagnosis: vaccine: KW

KW cytostatic; gene therapy; metastasis.
VZ

OS
Homo sapiens.

PN WG200151633-A2.

PD 19-JUL-2001.

1.6; JAN-2001; 2001WO-US01574.

PR 14-JAN-2000; 2000US-0483672.

PA (CORI-) CORIXA CORP.

XX	Xu J.	Dillon DC.	Mitcham II.	Harlocker SL.	Jiang Y.	Reed SC.
XX						
PT						

Kałos MD, Fanger GR
Wang A, Meacham MT

XX
XX
WPT. 2001-125873/4E

XX XX

diagnosing, monitoring and treating prostate cancer in a patient and

XX
XX

XX
Page 222, Page 245pp, English
XX

the present invention describes polynucleotide sequences (I) which encode prostate-specific proteins (II). (I) and (II) have cytostatic activity, and can be used in vaccine production and gene therapy. (I), (II), antibodies to (II), fusion proteins comprising (II), and isolated T cells prepared using (I) or (II) are used treat cancer in a patient. (I) and the antibodies are also used in the detection of cancer in a patient. The cancer that is diagnosed or treated is particularly prostate cancer. (I) and (II) can be used in vaccines. The antibodies or (I) can be used for monitoring the progression of cancer in a patient. (I) and (II) can also be used to improve diagnostic and therapeutic methods for prostate cancer. They can indicate the level of metastasis as, well as the prostate volume. AAH93357 to AAH93844 and AAH01115 to AAH01318 represent polynucleotide and amino acid sequences used in the exemplification of the present invention.

Sequence 209 AA:

alignment_scores:		
Quality:	69.50	Length: 77
Ratio:	1.580	Gaps: 3
Percent Similarity:	57.143	Percent Identity: 29.870

alignment block:

```
alignment_block:
  UIS-09-049-696-8  x  AAM01295
```

align seq 1/1 to: 200 from: 1 to: 200

[illegible][illegible]

100

1. $\frac{1}{2}$ 2. $\frac{1}{3}$ 3. $\frac{1}{4}$ 4. $\frac{1}{5}$ 5. $\frac{1}{6}$ 6. $\frac{1}{7}$ 7. $\frac{1}{8}$ 8. $\frac{1}{9}$ 9. $\frac{1}{10}$ 10. $\frac{1}{11}$ 11. $\frac{1}{12}$ 12. $\frac{1}{13}$ 13. $\frac{1}{14}$ 14. $\frac{1}{15}$ 15. $\frac{1}{16}$ 16. $\frac{1}{17}$ 17. $\frac{1}{18}$ 18. $\frac{1}{19}$ 19. $\frac{1}{20}$ 20. $\frac{1}{21}$ 21. $\frac{1}{22}$ 22. $\frac{1}{23}$ 23. $\frac{1}{24}$ 24. $\frac{1}{25}$ 25. $\frac{1}{26}$ 26. $\frac{1}{27}$ 27. $\frac{1}{28}$ 28. $\frac{1}{29}$ 29. $\frac{1}{30}$ 30. $\frac{1}{31}$ 31. $\frac{1}{32}$ 32. $\frac{1}{33}$ 33. $\frac{1}{34}$ 34. $\frac{1}{35}$ 35. $\frac{1}{36}$ 36. $\frac{1}{37}$ 37. $\frac{1}{38}$ 38. $\frac{1}{39}$ 39. $\frac{1}{40}$ 40. $\frac{1}{41}$ 41. $\frac{1}{42}$ 42. $\frac{1}{43}$ 43. $\frac{1}{44}$ 44. $\frac{1}{45}$ 45. $\frac{1}{46}$ 46. $\frac{1}{47}$ 47. $\frac{1}{48}$ 48. $\frac{1}{49}$ 49. $\frac{1}{50}$ 50. $\frac{1}{51}$ 51. $\frac{1}{52}$ 52. $\frac{1}{53}$ 53. $\frac{1}{54}$ 54. $\frac{1}{55}$ 55. $\frac{1}{56}$ 56. $\frac{1}{57}$ 57. $\frac{1}{58}$ 58. $\frac{1}{59}$ 59. $\frac{1}{60}$ 60. $\frac{1}{61}$ 61. $\frac{1}{62}$ 62. $\frac{1}{63}$ 63. $\frac{1}{64}$ 64. $\frac{1}{65}$ 65. $\frac{1}{66}$ 66. $\frac{1}{67}$ 67. $\frac{1}{68}$ 68. $\frac{1}{69}$ 69. $\frac{1}{70}$ 70. $\frac{1}{71}$ 71. $\frac{1}{72}$ 72. $\frac{1}{73}$ 73. $\frac{1}{74}$ 74. $\frac{1}{75}$ 75. $\frac{1}{76}$ 76. $\frac{1}{77}$ 77. $\frac{1}{78}$ 78. $\frac{1}{79}$ 79. $\frac{1}{80}$ 80. $\frac{1}{81}$ 81. $\frac{1}{82}$ 82. $\frac{1}{83}$ 83. $\frac{1}{84}$ 84. $\frac{1}{85}$ 85. $\frac{1}{86}$ 86. $\frac{1}{87}$ 87. $\frac{1}{88}$ 88. $\frac{1}{89}$ 89. $\frac{1}{90}$ 90. $\frac{1}{91}$ 91. $\frac{1}{92}$ 92. $\frac{1}{93}$ 93. $\frac{1}{94}$ 94. $\frac{1}{95}$ 95. $\frac{1}{96}$ 96. $\frac{1}{97}$ 97. $\frac{1}{98}$ 98. $\frac{1}{99}$ 99. $\frac{1}{100}$ 100. $\frac{1}{101}$ 101. $\frac{1}{102}$ 102. $\frac{1}{103}$ 103. $\frac{1}{104}$ 104. $\frac{1}{105}$ 105. $\frac{1}{106}$ 106. $\frac{1}{107}$ 107. $\frac{1}{108}$ 108. $\frac{1}{109}$ 109. $\frac{1}{110}$ 110. $\frac{1}{111}$ 111. $\frac{1}{112}$ 112. $\frac{1}{113}$ 113. $\frac{1}{114}$ 114. $\frac{1}{115}$ 115. $\frac{1}{116}$ 116. $\frac{1}{117}$ 117. $\frac{1}{118}$ 118. $\frac{1}{119}$ 119. $\frac{1}{120}$ 120. $\frac{1}{121}$ 121. $\frac{1}{122}$ 122. $\frac{1}{123}$ 123. $\frac{1}{124}$ 124. $\frac{1}{125}$ 125. $\frac{1}{126}$ 126. $\frac{1}{127}$ 127. $\frac{1}{128}$ 128. $\frac{1}{129}$ 129. $\frac{1}{130}$ 130. $\frac{1}{131}$ 131. $\frac{1}{132}$ 132. $\frac{1}{133}$ 133. $\frac{1}{134}$ 134. $\frac{1}{135}$ 135. $\frac{1}{136}$ 136. $\frac{1}{137}$ 137. $\frac{1}{138}$ 138. $\frac{1}{139}$ 139. $\frac{1}{140}$ 140. $\frac{1}{141}$ 141. $\frac{1}{142}$ 142. $\frac{1}{143}$ 143. $\frac{1}{144}$ 144. $\frac{1}{145}$ 145. $\frac{1}{146}$ 146. $\frac{1}{147}$ 147. $\frac{1}{148}$ 148. $\frac{1}{149}$ 149. $\frac{1}{150}$ 150. $\frac{1}{151}$ 151. $\frac{1}{152}$ 152. $\frac{1}{153}$ 153. $\frac{1}{154}$ 154. $\frac{1}{155}$ 155. $\frac{1}{156}$ 156. $\frac{1}{157}$ 157. $\frac{1}{158}$ 158. $\frac{1}{159}$ 159. $\frac{1}{160}$ 160. $\frac{1}{161}$ 161. $\frac{1}{162}$ 162. $\frac{1}{163}$ 163. $\frac{1}{164}$ 164. $\frac{1}{165}$ 165. $\frac{1}{166}$ 166. $\frac{1}{167}$ 167. $\frac{1}{168}$ 168. $\frac{1}{169}$ 169. $\frac{1}{170}$ 170. $\frac{1}{171}$ 171. $\frac{1}{172}$ 172. $\frac{1}{173}$ 173. $\frac{1}{174}$ 174. $\frac{1}{175}$ 175. $\frac{1}{176}$ 176. $\frac{1}{177}$ 177. $\frac{1}{178}$ 178. $\frac{1}{179}$ 179. $\frac{1}{180}$ 180. $\frac{1}{181}$ 181. $\frac{1}{182}$ 182. $\frac{1}{183}$ 183. $\frac{1}{184}$ 184. $\frac{1}{185}$ 185. $\frac{1}{186}$ 186. $\frac{1}{187}$ 187. $\frac{1}{188}$ 188. $\frac{1}{189}$ 189. $\frac{1}{190}$ 190. $\frac{1}{191}$ 191. $\frac{1}{192}$ 192. $\frac{1}{193}$ 193. $\frac{1}{194}$ 194. $\frac{1}{195}$ 195. $\frac{1}{196}$ 196. $\frac{1}{197}$ 197. $\frac{1}{198}$ 198. $\frac{1}{199}$ 199. $\frac{1}{200}$ 200. $\frac{1}{201}$ 201. $\frac{1}{202}$ 202. $\frac{1}{203}$ 203. $\frac{1}{204}$ 204. $\frac{1}{205}$ 205. $\frac{1}{206}$ 206. $\frac{1}{207}$ 207. $\frac{1}{208}$ 208. $\frac{1}{209}$ 209. $\frac{1}{210}$ 210. $\frac{1}{211}$ 211. $\frac{1}{212}$ 212. $\frac{1}{213}$ 213. $\frac{1}{214}$ 214. $\frac{1}{215}$ 215. $\frac{1}{216}$ 216. $\frac{1}{217}$ 217. $\frac{1}{218}$ 218. $\frac{1}{219}$ 219. $\frac{1}{220}$ 220. $\frac{1}{221}$ 221. $\frac{1}{222}$ 222. $\frac{1}{223}$ 223. $\frac{1}{224}$ 224. $\frac{1}{225}$ 225. $\frac{1}{226}$ 226. $\frac{1}{227}$ 227. $\frac{1}{228}$ 228. $\frac{1}{229}$ 229. $\frac{1}{230}$ 230. $\frac{1}{231}$ 231. $\frac{1}{232}$ 232. $\frac{1}{233}$ 233. $\frac{1}{234}$ 234. $\frac{1}{235}$ 235. $\frac{1}{236}$ 236. $\frac{1}{237}$ 237. $\frac{1}{238}$ 238. $\frac{1}{239}$ 239. $\frac{1}{240}$ 240

us-09-049-696-8.rag

Tue Apr 2 09:40:13 2002

82 slysAlaLeu.CysIleThrLeuThrLeuGlyThrPheLeuValGlyAla 98
111 GTTCAGAACAAATGGCCTCATTGATGCTTTT...GGGGCCCTTTTCATCAGG 157
:::|||||::: ||| |||::: |||:::
99 AlaLeuAlaAlaGlyLeuLeuTrpLysPheMetGlySerLysCysSerAs 115
158 AAATGGAGCTGTCTCTCAGCGCTCCATCCAGCTTGACAGTAAGGATTAA 207
:::|||| |::: |||::: |||::: |||
115 nSerGly.....IleGluCysAspSerSerGlyThrC 126
208 CCTCCAGAACACCCAGTGGATGAATGGC 236
:::|||| |::: |||::: |||::: |||
126 ysIleAsnProSerAsnTrpCysAspGly 135

OM of: US-09-049-696-13 to: A_Geneseq_1101.* out_format : pfs
Date: Mar 30, 2002 2:18 AM
About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+2p.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09049696/runat_28032002.145237.2001/app_query.fasta_1.12579
-DB=A_Geneseq_1101 -OFMT=fastan -SUFFIX=rag -GAPOP=12.000
-CAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-XGAPOP=6.000 -FGAPEXT=7.000 -TGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DICALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGNI_1_293 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-13
Query length: 227
Database: A_Geneseq_1101.*
Database sequences: 522463
Database length: 74073290
Search time (sec): 819.910000

score_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
/SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT:AAW06548	+		403.00	901.77	228	1.3e-42
/SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW46879	+		403.00	901.77	228	1.3e-42
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW73854	+		403.00	893.20	552	1.6e-42
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW75614	+		403.00	888.81	869	1.8e-42
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW74824	+		403.00	888.32	914	1.8e-42
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW42514	+		403.00	888.32	914	1.8e-42
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW73716	+		403.00	888.32	914	1.8e-42
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW74733	+		403.00	888.32	914	1.8e-42
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW74822	+		403.00	888.78	913	1.8e-42
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW73715	+		227.50	488.78	913	3.3e-20
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW166749	+		209.50	447.74	919	6.3e-18
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW87360	+		209.50	447.74	919	6.3e-18
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW65272	+		209.50	447.74	919	6.3e-18
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW45904	+		123.50	254.03	742	4.8e-07
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW41040	+		123.50	253.41	791	4.9e-07
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW11325	+		123.50	253.41	791	4.9e-07
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW70459	+		123.50	251.71	942	5.1e-07
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW41036	+		123.50	251.70	943	5.1e-07
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW11321	+		123.50	251.70	943	5.1e-07
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW74823	+		121.50	247.15	943	9.1e-07
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW52028	+		74.00	164.21	70	0.5117
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW03206	+		66.00	130.21	357	7.86
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW03208	+		66.00	130.21	357	7.86
/SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT:AAW13574	+		65.00	121.14	720	12.47
/SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT:AAW13573	+		65.00	120.27	787	12.75
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW06018	+		64.50	127.95	317	11.83
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW49277	+		64.50	127.95	317	11.83
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW27537	+		64.50	127.68	326	11.91
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW32083	+		64.50	127.68	326	11.91
/SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW59050	+		64.50	116.95	986	15.58
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW84992	+		64.50	103.02	4152	22.09
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW92746	+		64.00	124.85	388	14.38
/SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT:AAW48729	+		63.00	123.97	336	18.59
/SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT:AAW02701	+		63.00	123.97	336	18.59
/SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:AAW75704	+		63.00	113.89	951	23.93
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW84992	+		62.50	98.47	4152	39.60
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW22046	+		61.50	124.66	220	25.99
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW22045	+		61.50	123.42	250	26.81
/SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT:AAW27160	+		61.50	105.60	1572	41.89

/SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT:AAW27161 - 61.50 100.34 47.80
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW39043 + 61.50 95.16 54.42
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW18528 + 61.00 126.73 27.75
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW47147 + 61.00 123.70 29.94
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW44228 + 61.00 117.25 35.19
seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT:AAW06548
seq_documentation_block:
ID AAW06548 standard; Protein; 228 AA.
XX AAW06548;
AC AAW06548;
XX
DT 13-MAR-1997 (first entry)
XX
DE Human colon specific gene CSG5 polypeptide fragment.
XX
KW Colon specific gene: CSG5; colon cancer; metastasis; diagnosis;
KW therapy; antibody; vaccine.
XX
OS Homo sapiens.
XX
PN W09639419-A1.
XX
PD 12-DEC-1996.
XX
PF 06-JUN-1995; 95WO-US07289.
XX
PR 06-JUN-1995; 95WO-US07289.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Yu G;
XX
DR WPT: 1997-043054/04.
DR N-PSDB; AAT45884.
XX
PT Human colon specific genes and their expression products - detection
PT of which, in non-colon tissue samples, can be used as indication of
PT colon cancer metastasis
XX
PS Claim 8; Fig 5; 60pp; English.
XX
CC Novel polypeptides (AAW06545-53) are encoded by cDNA clones (see also
CC AAW45880-92) corresponding to 13 human colon specific genes,
CC designated CSG1, CSG2, etc., that are primarily expressed in
CC tissues derived from the colon. Recombinant CSG polypeptides can
CC be produced in transformed host cells. They are useful diagnostic
CC markers for colon cancer and for colon cancer metastasis and can
CC also be used to screen for (ant)agonist cpds. of therapeutic or
CC diagnostic value. Antibodies raised against the colon-specific
CC polypeptides may be used to target colon cancer cells or as part
XX of a colon cancer vaccine.
XX
SQ Sequence 228 AA;
alignment_scores:
Quality: 403.00 Length: 75
Ratio: 5.373 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-049-696-13 x AAW06548 ..
Align seq 1/1 to: AAW06548 from: 1 to: 228
1 GTCCGGCTCTGGAGAGTAAACGCCAGCCAGAGTATACCCCA 50
|||||
19 ValArgAlaLeuGlyGlyValAsnAlaAlaArgArgValIleProGI 35
'51 GCAGAGTGGAGACTGTACATACCTGCCTGGATTGAGATGATGAATAC 100
|||||

CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 552 AA;

alignment_scores: Quality: 403.00 Length: 75
Ratio: 5.373 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-13 x AAG73854 ..

Align seg 1/1 to: AAG73854 from: 1 to: 552

1 GTCCGGGCTCTGGAGGAGTTAACGCAGCCAGCAGGAGAGTGATACCCCA 50
|||||
313 ValArgAlaLeuGlyGlyValAlaAlaArgArgValIleProG1 329

51 GCAGAGTGGAGCACTGTACATACCTGGCTGGATTGAGATGATGAATAC 100
|||||
329 nGlnSerGlyAlaLeuTyrlleProGlyTrpIleGluAsnAspGluIleG 346

101 AATGGAATCCACCAAGACCTGAAATTAATAGGATGATGTTCAACACAAG 150
|||||
346 lntRpsnProProArgProGluIleAsnLysAspValGlnHisLys 362

151 CAAGTGTGTTTCAGCAACATCCTCGGAGGCTCATTTGTGGCTTCTGA 200
|||||
363 GlnValCysPheSerArgThrSerSerGlyGlySerPheValAlaSerAs 379

201 TGTCCCAATGCTCCCATACCTGAT 225

379 pValProAsnAlaProIleProAsp 387

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAG75614

seq_documentation_block:

ID AAG75614 standard; Protein; 869 AA.

XX AC AAG75614;

XX DT 03-SEP-2001 (first entry)

XX DE Human colon cancer antigen protein SEQ ID NO:6378.

XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX KW colorectal carcinoma.

XX OS Homo sapiens.

XX PN WO200122920-A2.

XX PD 05-APR-2001.

XX PF 28-SEP-2000; 2000WO-US26524.

XX PR 29-SEP-1999; 99US-0157137.

XX PR 03-NOV-1999; 99US-0163280.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX DR WPI; 2001-235357/24.

XX DR N-PSDB; AAH35019.

XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX PT useful for preventing, diagnosing and/or treating colorectal cancers -

XX PS Claim 11; Page 7851-7854; 9803pp; English.

XX CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 869 AA;

alignment_scores:

Quality: 403.00 Length: 75
Ratio: 5.373 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-13 x AAG75614 ..

Align seg 1/1 to: AAG75614 from: 1 to: 869

1 GTCCGGGCTCTGGAGGAGTTAACGCAGCCAGCAGGAGAGTGATACCCCA 50
|||||
630 ValArgAlaLeuGlyGlyValAlaAlaArgArgValIleProG1 646

51 GCAGAGTGGAGCACTGTACATACCTGGCTGGATTGAGATGATGAATAC 100
|||||
646 nGlnSerGlyAlaLeuTyrlleProGlyTrpIleGluAsnAspGluIleG 663

101 AATGGAATCCACCAAGACCTGAAATTAATAGGATGATGTTCAACACAAG 150
|||||
663 lntRpsnProProArgProGluIleAsnLysAspValGlnHisLys 679

151 CAAGTGTGTTTCAGCAACATCCTCGGAGGCTCATTTGTGGCTTCTGA 200
|||||
680 GlnValCysPheSerArgThrSerSerGlyGlySerPheValAlaSerAs 696

201 TGTCCCAATGCTCCCATACCTGAT 225

696 pValProAsnAlaProIleProAsp 704

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74824

seq_documentation_block:

ID AAB74824 standard; Protein; 914 AA.

XX AC AAB74824;

XX DT 13-JUN-2001 (first entry)

XX DE Human ICACC-1 protein sequence.

XX KW ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
XX KW interleukin 9 induced calcium activated chloride channel; IL-9;

XX KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
XX KW anti-inflammatory; immunomodulatory; cystic fibrosis;
XX KW inflammatory bowel disease; autoimmune disease.

XX OS Homo sapiens.

XX PN WO994620-A1.

PD 10-SEP-1999.
XX
PF 03-MAR-1999; 99WO-US04703.
XX
PR 03-MAR-1998; 98US-0076815.
XX
PA (MAGA-) MAGAININ PHARM INC.
XX
PI Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;
PI Nicolaides NC, Zhou Y, Dong Q;
XX
XX WPI: 1999-550979/46.
DR N-PSDB: AAF81927.
XX
XX New nucleic acid encoding calcium activated chloride channel, used to
PT identify, e.g. specific modulators for treating atopic allergy -
XX
XX Claim 11; Fig 4B; 75pp; English.
XX
XX The present sequence represents the human interleukin 9 (IL-9) induced
CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
CC have anti-allergic, anti-asthmatic, anti-inflammatory and
CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
CC used to alleviate asthma (or more generally atopic allergy), while those
CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
CC (increased levels) or IBD (reduced levels), also for monitoring
CC treatment of these conditions. The ICACC proteins can be used:
CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
CC reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
CC identify modulators and binding partners. ICACC polynucleotides can be
CC used to generate transgenic animals or recombinant cells, used to screen
CC for antagonists, also as a source of therapeutic antisense agents or
CC diagnostic probes (for quantifying mRNA expression, e.g. for
XX identification of modulators).
SD Sequence 914 AA;

alignment_scores:
Quality: 403.00 Length: 75
Ratio: 5.373 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-13 x AAB74824 ..

Align seg 1/1 to: AAB74824 from: 1 to: 914

1 GTGCGGCTCTGGAGGAGTTAAGCGAGCCAGACGAGAGTGATACCCCA 50
|||||
675 ValArgAlaLeuGlyValAsnAlaAlaArgArgValIleProGI 691
51 GCAGAGTGGAGCACTGACATACCTGGCTGGATTGAGATGATGAATAC 100
|||||
691 nGlnSerGlyAlaLeuTyrIleProGlyTrpIleGluAsnAspGluIleG 708
101 AATGGAATCCACCAAGACCTGAAATTAATAGGATGATGTTCAACACAAG 150
|||||
708 lnTrpAsnProProArgProGluIleAsnLysAspValGlnHisLys 724
151 CAAGTGTGTTTCAGCAACATCCTCGGGAGCGCTCAFTTGGCTTCTGCA 200
|||||
725 GlnValCysPheSerArgThrSerSerGlyGlySerPheValAlaSerAs 741
201 TGTCCCAATGCTCCCATACCTGAT 225
|||||
741 pValProAsnAlaProIleProAsp 749

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW24514

seq_documentation_block:
ID AAM24514 standard; Protein; 914 AA.
XX
AC AAM24514;
XX
DT 12-OCT-2001 (first entry)
XX
DE C902P predicted amino acid sequence.
XX
XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW immunogenic; gene therapy; vaccine; colonic cancer.
XX
OS Homo sapiens.
XX
PN WO200149716-A2.
PD 12-JUL-2001.
XX
XX 29-DEC-2000; 200WO-US35596.
PF
XX 30-DEC-1999; 99US-0476296.
PR 10-JAN-2000; 2000US-0480321.
PR 15-FEB-2000; 2000US-0504629.
PR 06-MAR-2000; 2000US-0519444.
PR 19-MAY-2000; 2000US-0573251.
PR 29-JUN-2000; 2000US-0609448.
PR 28-AUG-2000; 2000US-0649811.
XX
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
XX WPI: 2001-441847/47.
XX
XX Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer -
XX
XX Claim 2; Page 440-443; 472pp; English.

The present invention describes colon tumour associated proteins (I) and the polynucleotides (II) that encode them. (I) have cytostatic activity. (I) and (II) can be used in gene therapy and vaccine production. (I) and (II) may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate colon tumour associated protein (TCAP) expression, such as colonic cancer. For example, (I) and (II) may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of TCAPs by expressing inactive proteins or to supplement the patients own production of them. Additionally, (II) may be used to produce the TCAP proteins, by inserting the nucleic acids into a host cell culturing the cell to express the protein. (II) and its complementary sequences may also be used as DNA probes in diagnostic polymerase chain reaction (PCR) and hybridisation assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. (I) may also be used as antigens in the production of antibodies against TCAPs and in assays to identify modulators of TCAP expression and activity. Anti-(I) antibodies and antagonists may also be used to down regulate TCAP expression and activity. The anti-(I) antibodies may also be used as diagnostic agents for detecting the presence of TCAPs in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512 and AAM24494 to AAM24523 represent nucleotide and amino acid sequences given in the exemplification of the present invention.

Sequence 914 AA;

alignment_scores:
Quality: 403.00 Length: 75
Ratio: 5.373 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```
alignment_block:
US-09-049-696-13 x AAM24514 ..
Align seg 1/1 to: AAM24514 from: 1 to: 914

1 GTGCGGCTCTGGGAGGAGTTAAGCGAGCCAGCGAGAGTATACCCCA 50
675 ValArgAlaLeuGlyGlyValAlaAlaArgArgValIleProG1 691

51 GCAGAGTGGAGCACTGTACATACCTGGCTGGATTTGAGAAATCATCAACAA 100
691 nGlnSerGlyAlaLeuTyrIleProGlyTrpIleGluAsnAspGluIleG 708

101 AATGGAATCCACCAAGACCTGAAATTAATGAAGATGATGTTCAACACAAG 150
708 InTrpAsnProProArgProGluIleAsnLysAspAspValGlnHisLys 724

151 CAAGTGTGTTTCAGCAGACATCCTCGGGAGGCTCATTTGCTGGCTTCTGA 200
725 GlnValCysPheSerArgThrSerSerGlyGlySerPheValAlaSerAs 741

201 TGTCCCAATGCTCCCATACCTGAT 225
741 pValProAsnAlaProIleProAsp 749

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB73716
seq_documentation_block:
ID AAB73716 standard; Protein; 914 AA.
XX
AC AAB73716;
XX
DT 11-SEP-2001 (first entry)
XX
DE Human CLC1 protein, SEQ ID NO:2.
XX
KW Human CLC1; goblet cell; mouse Gob-5 orthologue; drug screening;
KW expression inhibition; antisense therapy; gene therapy;
KW chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 152 /note= "Encoded by AGG in AAH46124"
FT
XX
PN WO200138530-A1.
XX
PD 31-MAY-2001.
XX
PF 22-NOV-2000; 2000WO-JP08232.
XX
PR 24-NOV-1999; 99JP-0333479.
PR 27-APR-2000; 2000JP-0127589.
XX
PA (TAKE ) TAKEDA CHEM IND LTD.
XX
PI Nakanishi A, Morita S;
XX
WPI: 2001-355935/37.
DR N-PSDB; AAH46102, AAH46124.
XX
PT New antisense nucleotide, useful for treatment and prevention of
PT bronchial asthma and chronic obstructive pulmonary disease
XX
PS Claim 2; Page 76-80; 104pp; Japanese.
XX
CC The invention relates to an antisense nucleotide targetted to the mouse
CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
CC the CLC1 gene (coding sequence shown in AAH46102). The invention also
CC relates to an antibody specific for the Gob-5 protein, medical and
CC diagnostic compositions containing the antisense nucleotide or the
CC antibody, and methods and kits for screening for compounds which inhibit
```

```
the protein. Gob-5 and CLC1 are proteins expressed by goblet cells.
The antisense oligonucleotides and antibody are therefore useful for the
treatment and prevention of bronchial asthma and chronic obstructive
pulmonary disease. The present sequence represents human CLC1 protein.
XX
SQ Sequence 914 AA;
XX
alignment_scores:
Quality: 403.00 Length: 75
Ratio: 5.373 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
XX
alignment_block:
US-09-049-696-13 x AAB73716 ..
Align seg 1/1 to: AAB73716 from: 1 to: 914

1 GTGCGGCTCTGGGAGGAGTTAAGCGAGCCAGCGAGAGTATACCCCA 50
675 ValArgAlaLeuGlyGlyValAlaAlaArgArgValIleProG1 691

51 GCAGAGTGGAGCACTGTACATACCTGGCTGGATTTGAGAAATCATCAACAA 100
691 nGlnSerGlyAlaLeuTyrIleProGlyTrpIleGluAsnAspGluIleG 708

101 AATGGAATCCACCAAGACCTGAAATTAATGAAGATGATGTTCAACACAAG 150
708 InTrpAsnProProArgProGluIleAsnLysAspAspValGlnHisLys 724

151 CAAGTGTGTTTCAGCAGACATCCTCGGGAGGCTCATTTGCTGGCTTCTGA 200
725 GlnValCysPheSerArgThrSerSerGlyGlySerPheValAlaSerAs 741

201 TGTCCCAATGCTCCCATACCTGAT 225
741 pValProAsnAlaProIleProAsp 749

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB74733
seq_documentation_block:
ID AAB74733 standard; Protein; 914 AA.
XX
AC AAB74733;
XX
DT 12-JUN-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.
XX
KW Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
KW antialzheimer; antiparkinsonian; antimicrobial; vulnery; gene therapy;
KW immune disorder; hyperproliferative disorder; cardiovascular disease;
KW cancer; angiogenic disorder; neurological disorder; infectious disease;
KW wound healing; regeneration; chemotaxis; chromosome 1.
XX
OS Homo sapiens.
XX
WPI: 2001-355935/37.
DR N-PSDB; AAH46102, AAH46124.
XX
PT New antisense nucleotide, useful for treatment and prevention of
PT bronchial asthma and chronic obstructive pulmonary disease
XX
PS Claim 2; Page 76-80; 104pp; Japanese.
XX
CC The invention relates to an antisense nucleotide targetted to the mouse
CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
CC the CLC1 gene (coding sequence shown in AAH46102). The invention also
CC relates to an antibody specific for the Gob-5 protein, medical and
CC diagnostic compositions containing the antisense nucleotide or the
CC antibody, and methods and kits for screening for compounds which inhibit
```

DR WPI; 2001-147550/15.
DR N-PSDB; AAF81787.
XX
XX Nucleic acids encoding 25 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy
XX
XX
XX Claim 11; Page 459-460; 485pp; English.
PS
XX AAF81787 to AAF81817 encode the human secreted proteins given in AAF74733
CC to AAF81772. Human secreted proteins can have activities based on the
CC tissues and cells they are expressed in. Example of activities include:
CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;
CC antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;
CC vascular; anti-angiogenic; ophthalmological; neuroprotectant; and
CC anticonvulsant; antialzheimers; antiparkinsonian; antimicrobial; and
CC vaccine. Human secreted protein nucleotide sequences (NAMI) and proteins
CC (PEP1) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. For example, NAMI
CC and PEP1 may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patients genome
CC that affect the activity of proteins by expressing inactive proteins or
CC to supplement the patients own production of polypeptides. Disorders that
CC may be prevented, diagnosed and/or treated include immune disorders,
CC hyperproliferative disorders (e.g. cancers), cardiovascular diseases,
CC angiogenic disorders, neurological disorders, infectious diseases and/or
CC for promoting wound healing, regeneration and /or chemotaxis. AAF81778 to
CC AAF81786 and AAF74732 represent sequences used in the exemplification of
CC the present invention.
XX
XX Sequence 914 AA;
SQ

alignment_scores:
Quality: 403.00 Length: 75
Ratio: 5.373 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-13 x AAF74733 ..

Align seg 1/1 to: AAF74733 from: 1 to: 914

1 GTGCGGGCTCTGGAGGAGTTAACGCCAGCAGCGAGAGTGATACCCCA 50
675 ValArgAlaLeuGlyGlyValAlaAlaAlaArgArgValIleProG1 691
51 GCAGAGTGGACCTGTACATCTGCTGATTGAGATGAAATAC 100
691 nGlnSerGlyAlaLeuTyrlleProGlyTrpIleGluAsnAspGluIleG 708
101 AATGGAATCCACCAAGACCTGAATTAATAGGATGATGTTCAACACAAG 150
708 InTrpAsnProProArgProGluIleAsnLysAspAspValGlnHisLys 724
151 CAAGTGTCTTTCAGCAGAACATCTCGGAGGCTCATTGTGCTTCTGA 200
725 GlnValCysPheSerArgThrSerSerGlyGlySerPheValAlaSerAs 741
201 TGTCCTCCAAATGCTCCATACCTGAT 225
741 PvalProAsnAlaProIleProAsp 749

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAF75474

seq_documentation_block:
ID AAG75474 standard; protein; 925 AA.
XX
XX AAG75474;
XX
XX 03-SEP-2001 (first entry)
XX

Human colon cancer antigen protein SEQ ID NO:6238.
Human: colon cancer; colon cancer antigen; diagnosis; detection;
colorectal carcinoma; chromosome 1.
Homo sapiens.
WO200122920-A2.
05-APR-2001.
28-SEP-2000; 2000WO-US26524.
29-SEP-1999; 99US-0157137.
03-NOV-1999; 99US-0163280.
(HUMA-) HUMAN GENOME SCI INC.
Ruben SM, Barash SC, Birse CE, Rosen CA;
WPI; 2001-235357/24.
N-PSDB; AAF34879.
Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
useful for preventing, diagnosing and/or treating colorectal cancers -
Claim 11; Page 7686-7690; 9803pp; English.
AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
cancer-associated nucleic acid molecules (N) and proteins (P), where
the proteins are collectively known as colon cancer antigens. The colon
cancer antigens have cytostatic activity and can be used in gene
therapy and vaccine production. N and P may be used in the prevention,
diagnosis and treatment of diseases associated with inappropriate P
expression. For example, N and P may be used to treat disorders
associated with decreased expression by rectifying mutations or deletions
in a patient's genome that affect the activity of P by expressing
inactive proteins or to supplement the patients own production of P.
Additionally, N may be used to produce the colon cancer-associated Ps,
by inserting the nucleic acids into a host cell and culturing the cell
to express the proteins. N and P can be used in the prevention, diagnosis
and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
and AAB77789 represent sequences used in the exemplification of the
present invention.
N.B. Pages 666 to 682 and page 7053 of the sequence listing were
missing at time of publication, meaning no sequences are present for
SEQ ID NO:1027 to 1052, 7921 and 7922.

Sequence 925 AA;
alignment_scores:
Quality: 403.00 Length: 75
Ratio: 5.373 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-13 x AAG75474 ..

Align seg 1/1 to: AAG75474 from: 1 to: 925

1 GTGCGGGCTCTGGAGGAGTTAACGCCAGCAGCGAGAGTGATACCCCA 50
685 ValArgAlaLeuGlyGlyValAlaAlaAlaArgArgArgValIleProG1 702
51 GCAGAGTGGAGCAGCTGTACATCTGCTGATTGAGATGATGAATAC 100
702 nGlnSerGlyAlaLeuTyrlleProGlyTrpIleGluAsnAspGluIleG 719
101 AATGGAATCCACCAAGACCTGAATTAATAGGATGATGTTCAACACAAG 150
719 InTrpAsnProProArgProGluIleAsnLysAspValGlnHisLys 735

151 CAAGTGTGTTTCAGCAGAACATCTCGGAGGCTCATTTGTGCTTCTGA 200
|||||
736 GlnValCysPheSerArgThrSerGlyGlySerPheValAlaSerAs 752
201 TGTCCCAATGCTCCCATACCTGAT 225
|||||
752 pValProAsnAlaProIleProAsp 760

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74822

seq_documentation_block:
ID AAB74822 standard; Protein; 913 AA.

XX AAB74822;

XX 13-JUN-2001 (first entry)

XX Murine ICACC-1 protein sequence.

XX ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
interleukin 9 induced calcium activated chloride channel; IL-9;
calcium activated chloride channel; anti-allergic; anti-asthmatic;
anti-inflammatory; immunomodulatory; cystic fibrosis;
inflammatory bowel disease; autoimmune disease.

XX Mus sp.

XX WO9944620-A1.

XX 10-SEP-1999.

XX 03-MAR-1999; 99WO-US04703.

XX 03-MAR-1998; 98US-0076815.

XX (MAGA-) MAGAININ PHARM INC.

XX Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;

XX Nicolaides NC, Zhou Y, Dong Q;

XX WPI; 1999-550979/46.

XX N-PSDB; AAF81925.

XX New nucleic acid encoding calcium activated chloride channel, used to
identify, e.g. specific modulators for treating atopic allergy -

XX Claim 12; Fig 2; 75pp; English.

XX The present sequence represents the murine interleukin 9 (IL-9) induced
calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
have anti-allergic, anti-asthmatic, anti-inflammatory and
immunomodulatory activities. Compounds (A) that downregulate ICACC are
used to alleviate asthma (or more generally atopic allergy), while those
(B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
inflammatory bowel disease (IBD) (or other autoimmune diseases).
Measuring levels of ICACC-1 can be used in the diagnosis of asthma
(increased levels) or IBD (reduced levels), also for monitoring
treatment of these conditions. The ICACC proteins can be used:
(i) to raise specific antibodies (Ab); useful: (a) as immunoassay
reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
(A) to disrupt bonding between ICACC and its ligands; and (iii) to
identify modulators and binding partners. ICACC polynucleotides can be
used to generate transgenic animals or recombinant cells, used to screen
for antagonists, also as a source of therapeutic antisense agents or
diagnostic probes (for quantifying mRNA expression, e.g. for
identification of modulators).

XX Sequence 913 AA;

alignment_scores:
Quality: 227.50 Length: 74
Ratio: 3.792 Gaps: 2

Percent Similarity: 81.081 Percent Identity: 62.162
alignment_block:
US-09-049-696-13 x AAB74822 ..
Align seg 1/1 to: AAB74822 from: 1 to: 913
7 GCTCTGGGAGGAGTTAAGCAGCCAGCAGGAGAGTGATACCCCGACAGAG 56
|||||
678 AlaleuGlyGlyValThrSerAspArgGlnAlaAlaProProLysAs 694
57 TGGAGCAGCTGTACATACCTGGCTGGCTGGATGAGATGATGAATACAATGGA 106
|||||
694 nArgAlaMetTyrIleAspGlyTrpIleGluAspGlyGluValArgMetA 711
107 ATCCACCAAGACCTGAATTAATAAAGGATGATGTTCAACACACCAAGCAAGTG 156
|||||
711 snProArgProGluThrSerTyr.....ValGlnAspLysGlnIleu 725
157 TGTTCAGCAGAACATCTCGGAGGCTCATTTGTGCTTCTGTGATGCTCC 206
|||||
726 CysPheSerArgThrSerSerGlyGlySerPheValAlaThrAsnValPr 742
207 A...AATGCTCCCATACCTGAT 225
|||||
742 oAlaAlaAlaProIleProAsp 749

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB73715

seq_documentation_block:

ID AAB73715 standard; Protein; 913 AA.

XX AAB73715;

XX 11-SEP-2001 (first entry)

XX Mouse Gob-5 protein, SEQ ID NO:1.

XX Mouse Gob-5; murine; goblet cell; human CLCA1 orthologue; drug screening;
expression inhibition; antisense therapy; gene therapy; bronchial asthma;
chronic obstructive pulmonary disease; antiasthmatic.

XX Mus sp.

XX WO200138530-A1.

XX 31-MAY-2001.

XX 22-NOV-2000; 2000WO-JP08232.

XX 24-NOV-1999; 99JP-0333479.

XX 27-APR-2000; 2000JP-0127589.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Nakanishi A, Morita S;

XX WPI; 2001-355935/37.

XX N-PSDB; AAH46101, AAH46120.

XX New antisense nucleotide, useful for treatment and prevention of
bronchial asthma and chronic obstructive pulmonary disease -

XX Claim 1; Page 72-76; 104pp; Japanese.

XX The invention relates to an antisense nucleotide targetted to the mouse
Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
the CLCA1 gene (coding sequence shown in AAH46102). The invention also
relates to an antibody specific for the Gob-5 protein, medical and
diagnostic compositions containing the antisense nucleotide or the
antibody, and methods and kits for screening for compounds which inhibit
the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.
The antisense oligonucleotides and antibody are therefore useful for the

Tue Apr 2 09:39:41 2002

CC treatment and prevention of bronchial asthma and chronic obstructive
XX pulmonary disease. The present sequence represents mouse Gob-5 protein.
XX
SQ Sequence 913 AA;

alignment_scores:
Quality: 227.50 Length: 74
Ratio: 3.792 Gaps: 2
Percent Similarity: 81.081 Percent Identity: 62.162

alignment_block:
US-09-049-696-13 x AAB73715 ..

Align seg 1/1 to: AAB73715 from: 1 to: 913

```
7  GCTGTGGAGGAGTTAACGACGACGAGAGTATACCCAGCAGAG 56
|||||
678  AlalaLeuGlyValThrSerAspArgGlnArgAlaAlaProProIysAs 694
|||||
57  TGGAGCACTGTACATACCTGGCTGGATTGAGATGATGAATACATGGA 106
|||||
694  nArgAlaMetTyIleAspGlyTrpIleGluAspGlyGluValArgMeta 711
|||||
107  ATCCACCAAGACCTGAAATTAATAAGATGATGTTCAACACAGCAAGTG 156
|||||
711  snProProArgProGluThrSerTyr.....ValGlnAspIysGlnLeu 725
|||||
157  TGTTCAGCAGAACATCCTCGGAGGCTCATTTGTGGCTTCTGATGCC 206
|||||
726  CysPheSerArgThrSerSerGlyGlySerPheValAlaThrAsnValPr 742
|||||
207  A...AATGCTCCCATACCTGAT 225
|||||
742  oAlaAlaAlaProIleProAsp 749
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA56749

seq_documentation_block:

ID AAY66749 standard; protein; 919 AA.

XX AC AAY66749;
XX
XX
XX
XX 05-APR-2000 (first entry)
XX DE Membrane-bound protein PRO1124.
XX
XX Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW pharmaceutical; receptor immunoadhesin; gene mapping.
XX
XX Homo sapiens.
XX
XX WO9963088-A2.
XX
XX 09-DEC-1999.
XX
XX 02-JUN-1999; 99WO-US12252.
XX PF
XX
XX 02-JUN-1998; 98US-0087607.
XX PR 02-JUN-1998; 98US-0087609.
XX PR 02-JUN-1998; 98US-0087759.
XX PR 03-JUN-1998; 98US-0087827.
XX PR 04-JUN-1998; 98US-0088021.
XX PR 04-JUN-1998; 98US-0088025.
XX PR 04-JUN-1998; 98US-0088028.
XX PR 04-JUN-1998; 98US-0088029.
XX PR 04-JUN-1998; 98US-0088030.
XX PR 04-JUN-1998; 98US-0088033.
XX PR 04-JUN-1998; 98US-0088326.
XX PR 05-JUN-1998; 98US-0088167.
XX PR 05-JUN-1998; 98US-0088202.
XX PR 05-JUN-1998; 98US-0088212.
XX PR 05-JUN-1998; 98US-0088217.

PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 23-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.

PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
DR WPI; 2001-183260/18.
DR N-PSDB; AAF92092.

XX Eighty four nucleic acids encoding PRO polypeptides, useful in
PT molecular biology, including use as hybridization probes, and in
PT chromosome and gene mapping. -

XX Claim 12; Fig 70; 278pp; English.

XX The present sequence is a human PRO polypeptide (secreted and
CC transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
CC anti-PRO antibodies are useful for preparation of a medicament useful in
CC the treatment of a condition which is responsive to the PRO protein,
CC agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
CC employed as molecular weight markers for protein electrophoresis. The PRO
CC coding sequence has applications in molecular biology, including use as
CC hybridisation probes, and in chromosome and gene mapping.

XX Sequence 919 AA;

alignment_scores:
Quality: 209.50 Length: 75
Ratio: 3.612 Gaps: 1
Percent Similarity: 77.333 Percent Identity: 56.000

alignment_block:

US-09-049-696-13 x AAB87560 ..

Align seg 1/1 to: AAB87560 from: 1 to: 919

1 GTGCGGCTCTGGAGAGTACCGAGCCAGCAGGAGAGTATACCCCA 50

678 ValArgAlaHisGlyGlyAlaAsnThrAlaArgLeuLysLeuArgProPr 694

51 GCAGAGTGGAGCACTGTACACTCTGCTGATGAGATGATGAATAC 100

694 oLeuAsnArgAlaAlaTyrileProGlyTrpValValAsnGlyGluileg 711

101 AATGGAATCCACCAAGACCTGAATTAATGAATGATGATTCACACAG 150

711 luAlaAsnProProArgProGluileasp...GluAspThrGlnThr 726

151 CAAGTGTGTTTCAGCAACATCTCGGAGGCTCATTTGTGCTCTCTGA 200

727 LeuGluaspPheSerArgThrAlaSerGlyGlyAlaPheValSerGI 743

201 TGTCCTCAATGCTCCCATCTGAT 225

743 nValProSerLeuProLeuProAsp 751

seq_name: /SDS2/gcdata/geneseq/geneseq/AA2001.DAT.AAB65272

seq_documentation_block:

ID AAB65272 standard; Protein: 919 AA.

XX AC AAB65272;

XX 02-APR-2001 (first entry)

XX Human PRO1124 (UN0562) protein sequence SEQ ID NO:379.

XX Human; secreted and transmembrane protein; PRO; cytostatic;
KW cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW diagnostic assay.

XX Homo sapiens.

XX WO200073454-A1.

XX

PD 07-DEC-2000.

XX 30-MAR-2000; 2000WO-US08439.

XX 02-JUN-1999; 99WO-US12252.

XX 23-JUN-1999; 99WO-US141037.

XX 07-JUL-1999; 99US-0143048.

XX 26-JUL-1999; 99US-0144758.

XX 26-JUL-1999; 99US-0145698.

XX 28-JUL-1999; 99US-0146222.

XX 17-AUG-1999; 99US-0149396.

XX 15-SEP-1999; 99WO-US21090.

XX 15-SEP-1999; 99WO-US21547.

XX 08-OCT-1999; 99US-0158663.

XX 30-NOV-1999; 99WO-US28313.

XX 01-DEC-1999; 99WO-US28301.

XX 16-DEC-1999; 99WO-US30095.

XX 20-DEC-1999; 99WO-US30911.

XX 05-JAN-2000; 2000WO-US00219.

XX 06-JAN-2000; 2000WO-US00376.

XX 11-FEB-2000; 2000WO-US03565.

XX 18-FEB-2000; 2000WO-US04341.

XX 22-FEB-2000; 2000WO-US04414.

XX 24-FEB-2000; 2000WO-US04914.

XX 02-MAR-2000; 2000WO-US05004.

XX 15-MAR-2000; 2000WO-US05841.

XX 20-MAR-2000; 2000WO-US06884.

XX 20-MAR-2000; 2000WO-US07377.

XX (GETH) GENENTECH INC.

XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;

XX Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;

XX Grimaldi CJ, Gurney AL, Kijavini IJ, Napier MA, Pan J, Paoni NF;

XX Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;

XX Zhang Z;

XX WPI; 2001-032160/04.

XX N-PSDB; AA44241.

XX PRO polynucleotides used to produce polypeptides used to target

XX bioactive molecules such as toxins, radiolabels or antibodies, to

XX specific cells, to cause targeted cell death -

XX Claim 12; Fig 274; 935pp; English.

XX The present invention describes human secreted and transmembrane PRO
XX proteins. The PRO proteins have cytostatic activity. The PRO proteins
XX can be used for targeted delivery of bioactive molecules, such as
XX toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
XX sequences, and their fragments, can be used as hybridisation probes, in
XX chromosomal and gene mapping, and in the generation of anti-sense RNA
XX and DNA. They may also be used to produce transgenic animals which are
XX used to develop and screen therapeutically useful reagents. The PRO
XX nucleotide and protein sequence can be used for tissue typing and in
XX treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
XX AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
XX in the isolation of human PRO sequences. AAF44087 to AAF44269 and
XX AAF465154 to AAF65300 represent human PRO polynucleotide and protein
XX sequences given in the exemplification of the present invention.

alignment_scores:

Quality: 209.50 Length: 75

Ratio: 3.612 Gaps: 1

Percent Similarity: 77.333 Percent Identity: 56.000

alignment_block:

US-09-049-696-13 x AAB65272 ..

Align seg 1/1 to: AAB65272 from: 1 to: 919

```
1 GTCCGGCTCTGGAGAGTTAAGCCAGCCAGAGAGAGTATACCCCA 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
678 ValArgAlaHisGlyGlyAlaAsnThrAlaArgLeuLysLeuArgProPr 694
51 GCACAGTGGAGCACTGTACATACCTGCTGGATTGAGAATGATCAATAC 100
::: ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
694 oLeuAsnArgAlaAlaTyrIleProGlyTrpValValAsnGlyGluLeuG 711
101 AATGAATCCACCAAGACCTGAAATTAATAAGGATGATGTTCAACACAAG 150
:: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
711 luAlaAsnProProArgProGluIleAsp...GluAspThrGlnThrThr 726
151 CAAGTGTGTTTCACCAAGACATCTCCGGAGGCTCATTTGGGCTTCTGA 200
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
727 LeuGluAspPheSerArgThrAlaSerGlyGlyAlaPheValValSerG1 743
201 TGTCCCAATGCTCCATACCTGAT 225
|||||::: ||||| |||||
743 nValProSerLeuProLeuProAsp 751
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB45904

seq_documentation_block:

ID AAB45904 standard; Protein; 742 AA.

XX AAB45904;

XX 21-MAR-2001 (first entry)

XX Human tumor-associated antigen C42 protein.

XX Tumor-associated antigen; C42; human; immunogenic; cancer therapy;

KW cytostatic; immunotherapy; vaccine; lung; breast; esophagus.

XX Homo sapiens.

XX DE19924199-A1.

XX 30-NOV-2000.

XX 27-MAY-1999; 99DE-1024199.

XX 27-MAY-1999; 99DE-1024199.

XX (BOEH) BOEHRINGER INGELHEIM INT GMBH.

XX Adolf G, Heider K, Koenig U, Sommergruber W, Adolf GR, Heider K;

XX Koenig U, Sommergruber W;

XX WPI; 2001-062549/08.

XX N-PSDB; AAC82881.

XX Tumor-associated antigen C42, and related nucleic acids and antibodies,

XX useful in immunotherapy of cancer and for diagnosis -

XX Claim 1; Page 18; 40pp; German.

XX This invention describes a novel tumor-associated antigen C42 which has
CC cytostatic activity. The invention also describes (a) immunogenic protein
CC fragments or peptides (I) derived from C42; (b) a pharmaceutical
CC composition for parenteral, topical, oral or local administration
CC containing at least 1 C42 or (I); (c) an isolated DNA (II) encoding C42
CC or its fragments; (d) a recombinant DNA (III) that contains (II); and
CC (e) antibodies (Ab) raised against C42 or (I). C42 and its fragments
CC induce a humoral immune response and, when presented by major
CC histocompatibility complex molecules, a cellular immune response. C42,
CC and its fragments, are used for immunotherapy of cancers that express C42,
CC (particularly of lung, breast and esophagus), to raise specific
CC antibodies (Ab) and for diagnosis, e.g. to detect induction of an immune
CC response and for optimization of treatment. Ab are used therapeutically
CC (e.g. when conjugated to a cytotoxin) or for diagnosis or monitoring of
CC cancers that express C42. Nucleic acid (II) that encodes C42 can also be

CC used for immunotherapy and cells that express C42 as antitumor vaccines.

XX Sequence 742 AA;

alignment_scores:

Quality: 123.50 Length: 57

Ratio: 3.012 Gaps: 1

Percent similarity: 71.930 Percent identity: 45.614

alignment_block:

US-09-049-696-13 x AAB45904 ..

Align/Seg 1/1 to: AAB45904 from: 1 to: 742

155 AGTGGAGCACTGTACATACCTGCTGGATTGAGAATGATGAAATACAAATG 104

||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

506 SerHisAlaMetTyrValProGlyTyrThrAlaAsnGlyAsnIleGlnMe 522

||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

105 GAATCCACCAAGACCTGAAATTAATAAGGATGATGTTCAACACAAGCAAG 154

||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

522 tAsnAlaProArgLysSerValGlyArgAsnGluGluGluArgLysTrp. 538

155 TGTGTTTCACGACAGACATCTCCGGAGGCTCATTTGGGCTTCTGATCTC 204

||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

539 ..GlyPheSerArgValSerSerGlyGlySerPheSerValLeuGlyVal 554

||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

205 CCAATGCTCCCATACCTGAT 225

||| ||||| |||||

555 ProAlaGlyProHisProAsp 561

This Page Blank (uspto)


```
AC Q9UPC6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Eldie R.C., Ji H.L., Schreur K.D., Fuller C.M.,
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
RT of human CLCAL1, the first human member of the family of Ca2+-activated
RT Cl-channel proteins."
RL Genomics 54:200-214(1998).
RL EMBL; AF039401; AAC95429.1; -.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPase_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 914 AA; 100160 MW; 7761362B66752AA3 CRC64;

alignment_scores:
  Quality: 392.00      Length: 78
  Ratio: 5.026        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-12 x Q9UPC6 ..
Align seg 1/1 to: Q9UPC6 from: 1 to: 914

1 GACACAGCAAAATCCCGAGCCCTCTGGTAGTTATGCAAAATATTCGCCA 50
|||||
600 AspThrSerLysPheProSerProLeuValValThrAlaAsnIleArg61 616

51 AGGAGCCTCCCAATCTCAGGCCAGTGTACAGCCCTGATTAATCAG 100
|||||
616 nGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerV 633

101 TGAATGGAACAAACAGTTACCTTGGAACTACTGGATAATGGAGCAGTGCT 150
|||||
633 alaAsnGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAla 649

151 GATGCTACTAAGGATGACGGTGTCTACTCAAGTATTAATGCAACTATGCA 200
|||||
650 AspAlaThrLysAspAspGlyValThrSerArgTyrPheThrThrTyrAs 666

201 CACGAATGTAGATACAGTGTAAAGTGGCGGCT 234
|||||
666 pThrAsnGlyArgTyrSerValLysValArgAla 677

seq_name: sp_human:Q9UNF6

seq_documentation_block:
ID Q9UNF6 PRELIMINARY; PRT; 914 AA.
AC Q9UNF6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 1.
GN CACCL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE, AND COLON;
RX MEDLINE=99364503; PubMed=10437792;
RA Aguel M., Vermat T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
RT chloride channel (CaCC) family predominantly expressed in the
RT digestive tract and trachea."
RL FEBS Lett. 455:295-301(1999).
RL EMBL; AF127036; AAD25487.1; -.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPase_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
SQ SEQUENCE 914 AA; 100202 MW; B58B451D97887D6B CRC64;

alignment_scores:
  Quality: 392.00      Length: 78
  Ratio: 5.026        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-12 x Q9UNF6 ..
Align seg 1/1 to: Q9UNF6 from: 1 to: 914

1 GACACAGCAAAATCCCGAGCCCTCTGGTAGTTATGCAAAATATTCGCCA 50
|||||
600 AspThrSerLysPheProSerProLeuValValThrAlaAsnIleArg61 616

51 AGGAGCCTCCCAATCTCAGGCCAGTGTACAGCCCTGATTAATCAG 100
|||||
616 nGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerV 633

101 TGAATGGAACAAACAGTTACCTTGGAACTACTGGATAATGGAGCAGTGCT 150
|||||
633 alaAsnGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAla 649

151 GATGCTACTAAGGATGACGGTGTCTACTCAAGTATTAATGCAACTATGCA 200
|||||
650 AspAlaThrLysAspAspGlyValThrSerArgTyrPheThrThrTyrAs 666

201 CACGAATGTAGATACAGTGTAAAGTGGCGGCT 234
|||||
666 pThrAsnGlyArgTyrSerValLysValArgAla 677

seq_name: sp_rodent:O88826

seq_documentation_block:
ID O88826 PRELIMINARY; PRT; 913 AA.
AC O88826;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GC3-5 PROTEIN.
GN GC3-5 OR GOB-5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RA Komiya T., Tanigawa Y., Hirohashi S.;
RT "Cloning of a gene, gob-5, which is expressed in intestinal goblet
RT cells in mice."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB017156; BAA33743.1; -.
RL MGI; 1346342; Clca3.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
```



```
DR EMBL; AK008659; BAB25815.1; -.
DR MG; MGI; I346342; C1Ca3.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR02035; WFPA.
DR SMART; SM0032; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN1.
DR PROSITE; PS0234; WFPA; 1.
DR SEQUENCE 913 AA; 100111 MW; 491E584B06D9A89 CRC64;
```

alignment_scores:			
Quality:	339.00	Length:	78
Ratio:	4.403	Gaps:	0
Percent Similarity:	98.718	Percent Identity:	84.615

```
alignment_block:
US-09-049-696-12 x Q9D7Z6 ..
'Allian'seg 1/1 to: Q9D7Z6 from: 1 to: 913
```

1 GACACCAGCAAATTTCCCAGCCCTCTGGTAGTTTATGCAAAATATTGCCA 50

52 AGGAGCCTCCCAATTCTCAGGGCCAGTGTCACAGCCCTGATTGAATCAG 100

61.7 nGlyAlaSerProIleLeuArgAlaSerValThrAlaLeuIleGluSerV 634

634 aAsnGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAla 650

151 GATGCTACTAAGGATTGACGGTGTTCTACTCAAGGTATTTTCACAACCTTATGA 200
651 Acsn1JantherusAcenLenC1w1a1TurSorArvDhoDhombr1adhole 667

201 CACGAATGGTAGATACAGTGTAAAGTGGGGGCT 234

seq_name: sp_mammal:Q9TUB5
60/ PALAASHGLYARGITYRSERVALLYSIIETRPALA 6/8

```
seq_documentation_block:
ID Q9TUB5 PRELIMINARY; PRT; 917 AA.
```

DT	01-MAY-2000 (TREMBlrel. 13, Created)
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DE EPITHELIAL CHLORIDE CHANNEL PROTEIN.
GN AECC.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelele
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OY NCBI_TaxID=9823.

RN [1.]
RP SEQUENCE FROM N.A.
RC TISSUE=ILEAL MUCOSA:

RT "A. cDNA involved in porcine exocrine chloride conductance.
 RA Gašpar K.J., Gabriel S.E., Racette K.J., Forsyth G.W.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; vWFA.

DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWFA; 1.

100

Quality:	327.00	Length:	78
Ratio:	4.360	Gaps:	0

— 1997 —

alignment_block:

US-09-049-696-12 x Q9TUB5 ..

Align seg 1/1 to: Q9TUB5 from: 1 to: 917

```
1 GACACCAACAATCCCGAGCCCTCTGCTAGTATTGTCACAAATATTCGCCA 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
602 AspThrGlyLysPheProSerProMetValValThrLysIleHisC1 618
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 AGGAGCCTCCCAATTCACAGCCAGTGCACAGCCCTGATTGAATCAG 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
618 nGlyThrLeuProIleLeuArgAlaLysValThrAlaLeuIleGluSerG 635
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 TCAATGGAACAAACAGTACCTTGGAACTACTGGATAATGGAGCAGGTGCT 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
635 LuAsnGlyLysThrValThrLeuGluLeuLeuAspAsnGlyAlaGlyAla 651
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 GATGCTACTAGGATGAGCGGTGCTTACTCAAGGTATTTCACAACTTATGA 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
652 AspAlaThrLysAsnAspGlyIleTyrSerArgTyrPheThrAlaTyrAS 668
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
201 CACGAATGGTAGATACAGTGTAAAGTCCGGGCT 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
668 PalaAsnGlyArgTyrSerValLysValTrpAla 679
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

seq_name: sp_human:Q9NXP1

seq_documentation_block:

ID Q9NXP1 PRELIMINARY; PRT; 469 AA.

AC Q9NXP1;

DT 01-OCT-2000 (TRENBLrel. 15, Created)

DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)

DE CDNA FLJ20131 FIS, CLONE COL06357.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=COLON;

RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,

RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,

RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;

RT "NEDO human cDNA sequencing project.";

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK000138; BAA09096.1; -

SQ SEQUENCE 469 AA; 50694 MW; 6765986C886B3E56 CRC64;

alignment_scores:

Quality: 277.00

Ratio: 4.074

Percent Similarity: 87.179

Percent Identity: 69.231

alignment_block:

US-09-049-696-12 x Q9NXP1 ..

Align seg 1/1 to: Q9NXP1 from: 1 to: 469

```
1 GACACCAACAATCCCGAGCCCTCTGCTAGTATTGTCACAAATATTCGCCA 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
155 AspValAsnSerPheProSerProMetIleValTyrAlaGluIleLeuG1 171
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 AGGAGCCTCCCAATTCACAGCCAGTGCACAGCCCTGATTGAATCAG 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
171 nGlyThrValProValLeuGlyAlaAsnValThrAlaPheIleGluSerG 188
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 TCAATGGAACAAACAGTACCTTGGAACTACTGGATAATGGAGCAGGTGCT 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
188 LuAsnGlyHisThrGluValLeuGluLeuLeuAspAsnGlyAlaGlyAla 204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

seq_name: sp_rodent:Q9QX15

seq_documentation_block:

ID Q9QX15 PRELIMINARY; PRT; 902 AA.

```
151 GATGCTACTAGGATGAGCGGTGCTTACTCAAGGTATTTCACAACTTATGA 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
205 AspSerPheLysAsnAspGlyValTyrSerArgTyrPheThrAlaTyrTh 221
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

201 CACGAATGGTAGATACAGTGTAAAGTCCGGGCT 234

221 rGluAsnGlyArgTyrSerLeuLysValArgAla 232

seq_name: sp_human:Q9UNF7

seq_documentation_block:

ID Q9UNF7 PRELIMINARY; PRT; 917 AA.

AC Q9UNF7;

DT 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)

DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 2.

GN CAC2.

CS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=COLON;

RX MEDLINE=99364503; PubMed=10437792;

RA Aguel M., Vermet T., Culouscou J.M.;

RT "Identification of three novel members of the calcium-dependent

RT chloride channel (CaCC) family predominantly expressed in the

RT digestive tract and trachea";

RL FEBS Lett. 455:295-301(1999).

DR EMBL; AF127035; AAD48398.1; -

DR InterPro; IPR002035; VWFA.

DR Pfam; PF00092; vwa; 1.

DR PROSITE; PS50234; VWFA; 1.

DR SMART; SM00327; VWA; 1.

SQ SEQUENCE 917 AA; 101153 MW; EA01090D781BEB95 CRC64;

alignment_scores:

Quality: 277.00

Ratio: 4.074

Percent Similarity: 87.179

Percent Identity: 69.231

alignment_block:

US-09-049-696-12 x Q9UNF7 ..

Align seg 1/1 to: Q9UNF7 from: 1 to: 917

```
1 GACACCAACAATCCCGAGCCCTCTGCTAGTATTGTCACAAATATTCGCCA 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
603 AspValAsnSerPheProSerProMetIleValTyrAlaGluIleLeuG1 619
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 AGGAGCCTCCCAATTCACAGCCAGTGCACAGCCCTGATTGAATCAG 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
613 nGlyThrValProValLeuGlyAlaAsnValThrAlaPheIleGluSerG 636
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 TGAATGGAACAAACAGTACCTTGGAACTACTGGATAATGGAGCAGGTGCT 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
636 LuAsnGlyHisThrGluValLeuGluLeuLeuAspAsnGlyAlaGlyAla 652
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 GATGCTACTAGGATGAGCGGTGCTTACTCAAGGTATTTCACAACTTATGA 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
653 AspSerPheLysAsnAspGlyValTyrSerArgTyrPheThrAlaTyrTh 669
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

201 CACGAATGGTAGATACAGTGTAAAGTCCGGGCT 234

669 rGluAsnGlyArgTyrSerLeuLysValArgAla 680

seq_name: sp_rodent:Q9QX15

seq_documentation_block:

ID Q9QX15 PRELIMINARY; PRT; 902 AA.

```

AC Q90X15;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE CALCIUM-SENSITIVE CHLORIDE CONDUCTANCE PROTEIN-1.
OS CLCA1 OR MC1CA1.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=WHOLE LUNG;
RX MEDLINE=99041980; PubMed=9822685;
RA Gandhi R., Elble R.C., Gruber A.D., Schreur K.D., Ji H.-L.,
RA Fuller C.M., Pauli B.U.;
RT "Molecular and functional characterization of a calcium-sensitive
RT chloride channel from mouse lung.";
RL J. Biol. Chem. 273:32096-32101(1998).
DR EMBL; AF047838; AAC79982.1; -.
DR MGD; MGI:1316732; Clcal.
DR InterPro; IPR002035; VFWA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VFWA; 1.
SQ SEQUENCE 902 AA; 100107 MW; D9A915E698EA2D85 CRC64;

alignment_scores:
  Quality: 251.00      Length: 76
  Ratio: 3.922        Gaps: 0
Percent Similarity: 84.211 Percent Identity: 61.842

alignment_block:
US-09-049-696-12 x Q90X15
Align seg 1/1 to: Q90X15 from: 1 to: 902

  4 ACCAGCAAAATTCCTCCAGCCCTGCTAGTTATGCAAAATATTCGCCAAGG 53
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  606 ThrAlaGlnTyrProSerArgMetIleValTyrAlaArgValSerGlnG 622

  54 AGCCTCCCAATTCCTCAGGCCAGTGTCTACAGCCCTGATTGAATCAGTGA 103
  | |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  622 yPheLeuProValLeuGlyAlaAsnValThrAlaLeuIleGluAlaGluH 639

  104 ATGGAACAAACAGTACTTGGAACCTACTGGATAATGGACAGGTGCTGAT 153
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  639 isGlyHisGlnValThrLeuGluLeuTrpAspAsnGlyAlaGlyAlaAasp 655

  154 GCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCAACAACCTTATGACAC 203
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  655 ThrValIysAsnAspGlyIleTyrThrArgTyrPheThrAspTyrHisG 672

  204 GAATGTAGATACAGTGTAAAGTGGCGG 231
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  672 yAsnGlyArgTyrSerLeuIysValArg 681

seq_name: sp_rodent:088860

seq_documentation_block:
ID Q90X15 PRELIMINARY; PRT; 902 AA.
AC Q90X15;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE CALCIUM-SENSITIVE CHLORIDE CHANNEL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=WHOLE LUNG;
RX MEDLINE=99041980; PubMed=9822685;
RA Gandhi R., Elble R.C., Gruber A.D., Schreur K.D., Ji H.-L.,
RA Fuller C.M., Pauli B.U.;
RT "Molecular and functional characterization of a calcium-sensitive
RT chloride channel from mouse lung.";
RL J. Biol. Chem. 273:32096-32101(1998).
DR EMBL; AF047838; AAC79982.1; -.
DR MGD; MGI:1316732; Clcal.
DR InterPro; IPR002035; VFWA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VFWA; 1.
SQ SEQUENCE 902 AA; 100107 MW; D9A915E698EA2D85 CRC64;

alignment_scores:
  Quality: 251.00      Length: 76
  Ratio: 3.922        Gaps: 0
Percent Similarity: 84.211 Percent Identity: 61.842

alignment_block:
US-09-049-696-12 x Q90X15
Align seg 1/1 to: Q90X15 from: 1 to: 902

  4 ACCAGCAAAATTCCTCCAGCCCTGCTAGTTATGCAAAATATTCGCCAAGG 53
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  606 ThrAlaGlnTyrProSerArgMetIleValTyrAlaArgValSerGlnG 622

  54 AGCCTCCCAATTCCTCAGGCCAGTGTCTACAGCCCTGATTGAATCAGTGA 103
  | |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  622 yPheLeuProValLeuGlyAlaAsnValThrAlaLeuIleGluAlaGluH 639

  104 ATGGAACAAACAGTACTTGGAACCTACTGGATAATGGACAGGTGCTGAT 153
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  639 isGlyHisGlnValThrLeuGluLeuTrpAspAsnGlyAlaGlyAlaAasp 655

  154 GCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCAACAACCTTATGACAC 203
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  655 ThrValIysAsnAspGlyIleTyrThrArgTyrPheThrAspTyrHisG 672

  204 GAATGTAGATACAGTGTAAAGTGGCGG 231
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  672 yAsnGlyArgTyrSerLeuIysValArg 681

seq_name: sp_rodent:088860

seq_documentation_block:
ID Q90X15 PRELIMINARY; PRT; 901 AA.
AC Q90X15;
DT 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE CHLORIDE CHANNEL CACC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Romio L., Musante L., Cinti R., Moran O., Seri M., Galletta L.J.V.;
RT "Characterization of a murine gene homologous to the bovine CaCC
RT chloride channel.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF052746; AAC35003.1; -.
DR MGD; MGI:1316732; Clcal.
DR InterPro; IPR002035; VFWA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VFWA; 1.
SQ SEQUENCE 901 AA; 100038 MW; 13DF068BF3337AAB CRC64;

alignment_scores:
  Quality: 245.00      Length: 76
  Ratio: 3.828        Gaps: 0
Percent Similarity: 84.211 Percent Identity: 60.526

alignment_block:
US-09-049-696-12 x 088860
Align seg 1/1 to: 088860 from: 1 to: 901

  4 ACCAGCAAAATTCCTCCAGCCCTGCTAGTTATGCAAAATATTCGCCAAGG 53
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  606 ThrAlaGlnTyrProSerArgMetIleValTyrAlaArgValSerGlnG 622

  54 AGCCTCCCAATTCCTCAGGCCAGTGTCTACAGCCCTGATTGAATCAGTGA 103
  | |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  622 yPheLeuProValLeuGlyAlaAsnValThrAlaLeuIleGluAlaGluH 639

  104 ATGGAACAAACAGTACTTGGAACCTACTGGATAATGGACAGGTGCTGAT 153
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  639 isGlyHisGlnValThrLeuGluLeuTrpAspAsnGlyAlaGlyAlaAasp 655

  154 GCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCAACAACCTTATGACAC 203
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  655 ThrValIysAsnAspGlyIleTyrThrArgTyrPheThrAspTyrHisG 672

  204 GAATGTAGATACAGTGTAAAGTGGCGG 231
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  672 yAsnSerArgTyrSerLeuIysValArg 681

seq_name: sp_rodent:Q90X15

seq_documentation_block:
ID Q90X15 PRELIMINARY; PRT; 902 AA.
AC Q90X15;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE ENDOTHELIAL CHLORIDE CHANNEL.
GN CLCA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57/BL6;
RA Elble R.C., Pauli B.;
RT "A murine endothelial homolog of Lu-ECAM-1 associated with lung
RT melanoma metastasis.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF115852; AAG47626.1; -.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR002035; VFWA.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VFWA; 1.
SQ SEQUENCE 902 AA; 99855 MW; F2BA57DAA5D4AAAD CRC64;

alignment_scores:

```

[illegible]

OX	NCBI_TaxID=9913;	
RP	[1]	
RP	SEQUENCE FROM N.A.	
RP	TISSUE=LUNG;	
PC	Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,	
PA	Goodwin A., Pauli B.U.;	
RL	Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.	
DR	EMBL; AF001262; AAB86530.1; -	
DR	InterPro; IPR002035; VMFA.	
DR	PROSITE; PS0234; VMFA; 1.	
DR	SMART; SM00327; VWA; 1.	
SQ	SEQUENCE 794 AA; 88509 MW; B695E7256FC2C632 CRC64;	

```

alignment_scores:
  Quality: 243.00      Length: 77
  Ratio: 3.919        Gaps: 0
  Percent Similarity: 80.519      Percent Identity: 59.740

alignment_block:
  US-09-049-696-12 x O18742      ..

Align seg 1/1 to: O18742 from: 1 to: 794

```

675 yPheLeuProValLeuGlyLeSerValIleAlaIleIleGluThrGluA 642

104 ATGGAAAAACAGTTTACCTTGGAACTACTGGATAATGGACGAGGTGCTGAT 153

642 sPGlyHisGlnValThrLeuGluLeuTrpAsnNGlyAlaGlyArgAsp 658

154 GCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCACAACCTTATGACAC 203

659 ThrValIysAsnAspGlyIleTyrSerArgTyrPheThrAspTyrTyrG 675

204 GAATGCTGATACAGTGTAAAGTGGCGGCT 234

675 yAsnGlyArgTyrSerLeuIysValHisAla 685

204 GAATGCTAGATACACTTAAAACTGCGGGCT 234
 |||||
 675 yAsnGlyArgTyrSerLeuLysValHisAla 685

This Page Blank (uspto)

OM of: US-09-049-696-12 to: SwissProt_39:* out_format : pfs
 Date: Mar 30, 2002 2:52 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+ntp.model -DEV=xlp
 -O/cg2_1/USPTO.spool/US09049696/runat_28032002_145238_2085/app_query.fasta_1.12579
 -DB=SwissProt_39 -QFMT=fastan -SUFFIX=rs -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
 -TRN=human40.cdi -LIST=45 -DALIGN=200 -THR_SCORE=ptc
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=200000000
 -USER=US09049696 -CGNL_1_165 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPY
 -WAIT -THREADS=1

Search information block:
 Query: US-09-049-696-12
 Query length: 235
 Database: SwissProt_39:*
 Database sequences: 100059
 Database length: 36664827
 Search time (sec): 306.030000

score_list:	Strd	Orig	ZScore	EScore	Len	Documentation
SwissProt_39:ECIC_BOVIN	+	252.00	516.49	4.7e-22	903	! P54281 bos taurus (bovine).
SwissProt_39:PTSA_PEPDE	+	79.50	150.45	0.1606	651	! P43470 pedicoccus pentosace
SwissProt_39:RECN_MYCLE	-	68.50	127.80	3.26	587	! Q49896 mycobacterium leprae.
SwissProt_39:HYDH_ECOLI	+	67.50	127.59	4.22	465	! P14377 escherichia coli. sen
SwissProt_39:MYSL_YEAST	+	65.00	110.45	9.16	1928	! P08964 saccharomyces cerevi
SwissProt_39:FIG2_YEAST	+	63.00	107.68	15.68	1609	! P25653 saccharomyces cerevi
SwissProt_39:LPXK_NEIMA	+	62.50	119.40	16.31	344	! Q39544 neisseria meningitidi
SwissProt_39:LPXK_NEIMA	+	62.50	119.40	16.31	344	! Q39544 neisseria meningitidi
SwissProt_39:VFPI_BPP2	+	62.00	117.19	18.87	395	! P22501 bacteriophage p2. maj
SwissProt_39:ABCL_YEAST	+	62.00	115.21	19.15	501	! P27697 saccharomyces cerevis
SwissProt_39:CMF_DICDI	+	61.50	113.06	22.15	571	! P34090 dictyostelium discoide
SwissProt_39:EFIA_DICDI	+	61.00	113.86	25.05	456	! P18624 dictyostelium discoide
SwissProt_39:YF58_YEAST	+	61.00	109.64	25.86	758	! Q39299 saccharomyces cerevis
SwissProt_39:CBPA_CLOC	+	61.00	102.26	27.37	1848	! P38058 clostridium cellulov
SwissProt_39:VGL_BPFI	+	60.50	114.91	28.26	353	! O80299 bacteriophage ifi. mat
SwissProt_39:YCE_SCHPO	+	60.50	113.26	28.62	431	! Q09888 schizosaccharomyces r
SwissProt_39:NRG_DROME	+	60.50	104.09	30.70	1302	! P20241 drosophila melanogast
SwissProt_39:MLH3_HUMAN	+	60.50	103.18	30.91	1453	! Q9uhcl homo sapiens (human)
SwissProt_39:VIT_ONCMY	+	60.50	102.08	31.17	1659	! Q92093 oncorhynchus mykiss (
SwissProt_39:UCRI_SPIO	+	60.00	116.80	31.69	247	! P08980 spinacia oleracea (sp
SwissProt_39:CITA_KLEPN	+	60.00	110.21	33.33	547	! P52687 klebsiella pneumoniae
SwissProt_39:FATH_HUMAN	+	60.00	92.57	38.13	4590	! Q14517 homo sapiens (human)
SwissProt_39:EL3A_SOYEN	+	59.50	112.92	37.14	347	! Q33773 glycine max (soybean)
SwissProt_39:FAT_DROME	+	59.50	90.56	44.06	5147	! P33450 drosophila melanogast
SwissProt_39:N551_SOYEN	+	59.00	119.55	40.16	137	! Q05444 glycine max (soybean)
SwissProt_39:VM17_BORHE	+	59.00	111.70	42.64	353	! P32777 borrelia hermslii. var
SwissProt_39:YC78_MYCTU	+	59.00	104.18	45.16	875	! Q11042 mycobacterium tubercu
SwissProt_39:A180_RAT	+	59.00	103.81	45.29	915	! Q05140 rattus norvegicus (rat)
SwissProt_39:FASL_MOUSE	+	58.50	112.59	48.18	279	! P41047 mus musculus (mouse)
SwissProt_39:VGL_BPFI	+	58.50	110.75	48.86	348	! P03657 bacteriophage fl. gene
SwissProt_39:VGL_BPFD	+	58.50	110.75	48.86	348	! P03655 bacteriophage fl. gene
SwissProt_39:VGL_BPM13	+	58.50	110.75	48.86	348	! P03656 bacteriophage m13. gen
SwissProt_39:HS47_CHICK	+	58.50	109.50	49.33	405	! P13731 gallus gallus (chicken)
SwissProt_39:PAK5_HUMAN	+	58.50	104.74	51.16	719	! Q9P286 homo sapiens (human)
SwissProt_39:TSR_SHEEP	+	58.50	104.23	51.36	764	! P56495 ovis aries (sheep)
SwissProt_39:CPHA_ANASO	+	58.50	102.87	51.90	901	! Q86109 anasaeana sp. (strain
SwissProt_39:FL3H_MATIN	+	58.00	109.47	56.13	357	! Q05965 matthiola incana (com
SwissProt_39:FUMB_ECOLI	+	58.00	105.92	57.68	548	! P14407 escherichia coli. fum
SwissProt_39:ILVH_LACIA	+	57.50	115.16	61.14	158	! Q02140 lactococcus lactis (s
SwissProt_39:YF75_ARCTU	+	57.50	112.80	62.25	210	! Q28697 archaeglobus fulgidu

SwissProt_39:KRE9_YEAST + 57.50 110.54 63.34 276 ! P39005 saccharomyces cere
 SwissProt_39:TEHA_ECOLI + 57.50 109.06 64.06 330 ! P23396 escherichia coli.
 SwissProt_39:SHU7_ECOLI + 57.50 106.38 65.38 456 ! P09751 escherichia coli.
 SwissProt_39:MURF_BORBU + 57.50 106.23 65.46 464 ! Q44777 borrelia burgdorfe
 SwissProt_39:KDC2_DROME + 57.50 105.58 65.78 502 ! P16912 drosophila melanog
 seq_name: SwissProt_39:ECIC_BOVIN
 seq_documentation_block:
 ID ECIC_BOVIN STANDARD; PRT; 903 AA.
 AC P54281;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE EPIITHELIAL CHLORIDE CHANNEL PROTEIN (CALCIUM-ACTIVATED CHLORIDE
 CHANNEL).
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Trachea.
 RX MEDLINE=96125078; PubMed=8537359;
 RA Cunningham S.A., Awayda M.S., Buben J.K., Ismailov I.I.,
 RA Arrate M.P., Berdiev B.K., Benos D.J., Fuller C.M.;
 RT "Cloning of an epithelial chloride channel from bovine trachea.";
 RL J. Biol. Chem. 270:31016-31026(1995).
 CC !- FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE
 CC SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME;
 CC MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND
 CC TRANSEPITHELIAL TRANSPORT.
 CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC !- TISSUE SPECIFICITY: TRACHEA.
 CC !- PTM: PHOSPHORYLATED BY CAM-KINASE II.
 CC !- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U36445; AAC48511.1; -
 CC InterPro: IPR002035; VWFA.
 CC SMART: SM00327; VWFA. 1
 CC PROSITE: PS50234; VWFA; 1.
 CC Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
 CC Calcium channel; Phosphorylation; Glycoprotein.
 CC TRANSMEM 7 27 POTENTIAL.
 CC TRANSMEM 331 351 POTENTIAL.
 CC TRANSMEM 617 637 POTENTIAL.
 CC TRANSMEM 883 903 POTENTIAL.
 CC LOMAIN 308 476 VWFA.
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 75 75 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 360 360 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 504 504 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 688 688 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 811 811 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 816 816 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 842 842 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 857 857 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC SEQUENCE 903 AA; 100305 MW; 9742AB1590908AEC CRC64;

alignment_scores: Quality: 252.00 Length: 78

Ratio: 3.877 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 60.256
alignment_block:
US-09-049-696-12 x ECLC_BOVIN ..
Align seg 1/1 to: ECLC_BOVIN from: 1 to: 903
1 GACACGACCAATCCCGAGCCTCTGCTAGTTTATGCAAAATATTCGCCA 50
607 AsnThrAlaHisTyrProSerProValIleValTyrAlaGlnValSerG1 623
51 AGGAGCCTCCCAATCTCAGGCCAGTGTACAGCCCTGATTGAATCAG 100
623 nglyPheLeuProValLeuGlyIleAsnValThrAlaIleIleGluThrG 640
101 TGAATGGAAAAACAGTACCTTGGAACTACTGATAATGGAGCAGTGCT 150
640 luAspGlyHisGlnValThrLeuGluLeuTrpAspAsnGlyAlaGlyAla 656
151 GATGCTACTAAGATGACGGTCTCTACTCAAGGTATTTTCAACAATTCGA 200
657 AspThrValLysAsnAspGlyIleTyrSerArgTyrPheThrAspTyrAr 673
201 CACGAATGTAGATACAGTGTAAAGTGGGGCT 234
673 gGlyAsnGlyArgTyrSerLeuLysValHisAla 684
seq_name: SwissProt_39:PTSA_PEDPE

seq_documentation_block:
ID PTSA_PEDPE STANDARD; PRT; 651 AA.
AC P43470;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PTS SYSTEM, SUCROSE-SPECIFIC IIABC COMPONENT (EIIABC-SCR) (SUCROSE-
DE PERMEASE IIABC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, ABC
DE COMPONENT) (EC 2.7.1.69) (EII-SCR).
GN SCRA.
OS *Pediococcus pentosaceus*.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC *Pediococcus*.
OX NCBI_TaxID=1255;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PP1.0;
RA Leenhouts K.K.J., Bolhuis A.A., Kok J.J., Venema G.G.;
RL Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHONOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -!- CATALYTIC ACTIVITY: PROTEIN N-PHOSPHOISTIDINE + SUGAR =
CC PROTEIN HISTIDINE + SUGAR PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS A PTS EIIA DOMAIN.
CC -!- SIMILARITY: CONTAINS A PTS EIIIB DOMAIN.
CC -!- SIMILARITY: CONTAINS A PTS EIIIC DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 232771; CAA83668.1; -.
CC EMBL; L32093; AAA25567.1; -.
DR
DR

DR HSP; P20166; IAX3.
DR InterPro; IPR001127; PTS_EIIA.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00358; PTS_EIIA_1; 1.
DR Pfam; PF00387; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD001476; PTS_EIIB; 1.
DR ProDom; PD002243; PTS_EIIA; 1.
DR ProSITE; PS00371; PTS_EIIA_1; 1.
DR ProSITE; PS01035; PTS_EIIB_CYS; 1.
KW Phosphotransferase system; Sugar transport; Transferase;
KW Phosphorylation; Transmembrane.
FT DOMAIN 1 40 EIIB DOMAIN.
FT DOMAIN ? EIIIC DOMAIN.
FT DOMAIN 510 651 EIIA DOMAIN.
FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 324 324 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 562 562 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 651 AA; 68454 MW; C87BA09D550A77F8 CRC64;
alignment_scores:
Quality: 79.50 Length: 87
Ratio: 1.622 Gaps: 4
Percent Similarity: 56.322 Percent Identity: 29.885
alignment_block:
US-09-049-696-12 x PTSA_PEDPE ..
Align seg 1/1 to: PTSA_PEDPE from: 1 to: 651

1 GACACGACCAATCCCGAGCCTCTGCTAGTTTATGCAAAATATTCGCCA 50
475 AspGlnValLysSerProAlaProThrSerThrValIleAsnValasn.. 490
51 AGGAGCCTCCCAATCTCAGGCCAGTGTACAGCCCTGATTGAATCA. 99
491AspGluIleIleSerAlaProValThrGlyAlaSerGluSerL 505
100GTGAATGGAAAAACAGTTCCTGGAACACTGCTGTAATGGA 141
505 euLysGlnValAsnAspGlnValPheSerAlaGluIleMetGlyLysGly 521
142 GCAGTGCTGCTACTAAGGAT.....GACGGTGT 173
522 AlaAlaIleValProSerSerAspGlnValValAlaProAlaAspGlyVa 538
174 CTACTCAAGGTATTTTACAACTTATGACACGAATGCTAGATACAGTGAA 223
538 lIleThr.....ValThrTyrAspSerHisAlaTyrGlyIleU 552
224 AAGTCGGGGCT 234
552 ysThrThrAla 555

seq_name: SwissProt_39:RECN_MYCLE
seq_documentation_block:
ID RECN_MYCLE STANDARD; PRT; 587 AA.
AC Q49896; O08103;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE DNA REPAIR PROTEIN RECN (RECOMBINATION PROTEIN N).
GN RECN OR M1360 OR MLC1351.12C.
OS *Mycobacterium leprae*.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;


```
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=TN;
RC MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churchev C., Harris D.,
RA Mundall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feitwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Bartell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -1- FUNCTION: MAY BE INVOLVED IN RECOMBINATIONAL REPAIR OF DAMAGED
CC DNA (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE REC N FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U00021; AAA50914.1;
CC EMBL: 295117; CAB08275.1;
CC EMBL: AL583921; CAC31741.1;
CC Leptoma; M11360;
CC InterPro: IPR003439; ABC_transport.
CC DNA repair; ATP-binding; Complete proteome.
CC NP_BIND 29 36 ATP (POTENTIAL).
CC SEQUENCE 587 AA; 62723 MW; 58BC63F38A4DB732 CRC64;

alignment_scores:
  Quality: 68.50 Length: 46
  Ratio: 1.803 Gaps: 2
  Percent Similarity: 82.609 Percent Identity: 39.130

alignment_block:
  US-09-049-696-12/rev x REC_NMYCLE ..
  Align seg 1/1 to: REC_NMYCLE from: 1 to: 587
  158 GTACATCAGCACCTGCTCCATTATCCAGTAGTCCCAAGGTAACTGTTT 109
  259 MetGlycylAlaLysAlaAlaLeuGluSerThrAspAlaThrLeuAr 275
  108 TCCATTCAGTGATTCATC...AGGGGTGTGACACTGGCCCTGAGAATTG 62
  275 gAlaPheAlaAspGlnValGlyGluValLeuThrValValVal.GluVal 291
  61 GGGAGGCTCTCTGGCGAATATTCATAACTACCA 26
  292 GlyArgGluLeuGlyGluTyrLeuGluGluPro 303
  seq_name: SwissProt_39:HYDH_ECOLI

seq_documentation_block:
  ID HYDH_ECOLI STANDARD; PRT; 465 AA.
  AC P14377;
  DT 01-JAN-1990 (Rel. 13, Created)
  DT 01-OCT-1993 (Rel. 27, Last sequence update)
  DT 20-AUG-2001 (Rel. 40, Last annotation update)
  DE SENSOR PROTEIN HYDH (EC 2.7.3.-).
  GN HYDH OR B4003.
  OS Escherichia coli.
  OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
  OC Escherichia.
```

362 LeuAsnLeuTyrLeuAsnAlaIleGlnAlaIleGlyGlnHisGlyValI 378
75 CAGTGTACAGCCCTGATTGAATCAGTGAATGGAAGAAACAGTACCTTGG 124
|||||
378 eserValThrAla.....SerGluSerGlyAlaGlyValIleHis 392
125 AACTACTGATAATGACAGCAGTGTCTGATCTACTAGGATGACGGTGTCT 174
|||||
392 erValThrAspSerGlyGlyIleAlaAlaAspGlnLeuAspAlaIle 408
175 TACTCAAGTATTCACAACT 195
|||||
409 PheThrProTyrPheThrThr 415

seq_name: SwissProt_39.MYS1_YEAST

seq_documentation_block:

ID MYS1_YEAST STANDARD; PRT: 1928 AA.

AC P08964;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE MYOSIN-1 ISOFORM (TYPE II MYOSIN).

CN MYO1 OR YHR023W.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RC MEDLINE=91088308; PubMed=2263482;

RA Sweeney F.P., Watts F.Z., Pocklington M.J., Orr E.;

RA "The MYO1 gene from Saccharomyces cerevisiae: its complete nucleotide

sequence";

RL Nucleic Acids Res. 18:7147-7147(1990).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C / AB972;

RC MEDLINE=94378003; PubMed=8091229;

RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,

RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,

RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,

RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,

RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,

RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,

RA Vaudin M.;

RA "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome

VIII";

RL Science 265:2077-2082(1994).

RN [3]

RP SEQUENCE OF 1-760 FROM N.A.

RC STRAIN=S288C;

RC MEDLINE=88111539; PubMed=3322809;

RA Watts F.Z., Shields G., Orr E.;

RA "The yeast MYO1 gene encoding a myosin-like protein required for cell

division";

RL EMBO J. 6:3499-3505(1987).

CC -!- FUNCTION: REQUIRED FOR CELL DIVISION.

CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X53947; CAA37894.1; -

DR EMBL; X06187; CAA29550.1; -

DR EMBL; U10399; AAB68872.1; -

PIR; S05806; S05806.
PIR; S12323; S12323.
PIR; S46773; S46773.
HSP; P08799; 1MND.
SGD; S0001065; MYO1.
InterPro; IPR000048; IQ.
InterPro; IPR001609; myosin_head.
PIfam; PF00063; myosin_head; 1.
PRINTS; PR00193; MYOSINHEAVY.
ProDom; PD000355; myosin_head; 1.
SMART; SM00015; IQ; 1.
SMART; SM00242; MYSC; 1.
PROSITE; PS50096; IQ; 1.
Myosin; Actin-binding; ATP-binding; Coiled coil; Alkylation.
FT DOMAIN 1 793 MYOSIN HEAD-LIKE.
FT DOMAIN 794 823 IQ.
FT DOMAIN 856 1911 COILED COIL (POTENTIAL).
FT NE_BIND 180 187 ATP (BY SIMILARITY).
FT DOMAIN 460 529 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 703 703 ALKYLATION (BY SIMILARITY).
FT CONFLICT 36 36 K -> I (IN REF. 1).
FT CONFLICT 46 46 I -> T (IN REF. 1 AND 3).
FT CONFLICT 59 59 V -> S (IN REF. 1 AND 3).
FT CONFLICT 86 86 L -> F (IN REF. 1).
FT CONFLICT 330 330 MISSING (IN REF. 1 AND 3).
FT CONFLICT 343 343 N -> S (IN REF. 1 AND 3).
FT CONFLICT 421 426 QAKFI -> TKLSSL (IN REF. 1).
FT CONFLICT 515 515 D -> S (IN REF. 1).
FT CONFLICT 529 535 SKGPPTG -> ARGHDR (IN REF. 1 AND 3).
FT CONFLICT 541 541 D -> V (IN REF. 1 AND 3).
FT CONFLICT 550 551 TD -> LM (IN REF. 1).
FT CONFLICT 573 573 R -> A (IN REF. 1).
FT CONFLICT 582 582 H -> D (IN REF. 1).
FT CONFLICT 588 599 EYVEGWSLKNK -> NTLWKAGYPKT (IN REF. 1).
FT CONFLICT 599 599 MISSING (IN REF. 3).
FT CONFLICT 627 632 EKSSA -> GKNLLVC (IN REF. 1 AND 3).
FT CONFLICT 695 695 R -> S (IN REF. 1 AND 3).
FT CONFLICT 736 742 ENSTTTT -> RKNHHD (IN REF. 3).
FT CONFLICT 756 756 E -> R (IN REF. 1 AND 3).
FT CONFLICT 773 784 NTKLFFRAGVLA -> ILTVFKLEYWS (IN REF. 1).
FT CONFLICT 793 794 KL -> NV (IN REF. 1).
FT CONFLICT 896 896 N -> T (IN REF. 1).
FT CONFLICT 900 900 N -> NSQITKINTNITPTQSTYIGRPKRVCIGN (IN REF. 1).
FT CONFLICT 906 906 N -> I (IN REF. 1).
FT CONFLICT 911 911 N -> K (IN REF. 1).
FT CONFLICT 915 930 NESLNKVTSSSETLQ -> RIAIKLKPAINIT (IN REF. 1).
FT CONFLICT 934 939 DDLVSE -> MTLFL (IN REF. 1).
FT CONFLICT 951 953 AQN -> RKI (IN REF. 1).
FT CONFLICT 955 958 EEAH -> KKLD (IN REF. 1).
FT CONFLICT 1002 1002 S -> C (IN REF. 1).
FT CONFLICT 1049 1049 L -> D (IN REF. 1).
FT CONFLICT 1056 1056 C -> S (IN REF. 1).
FT CONFLICT 1060 1060 M -> I (IN REF. 1).
FT CONFLICT 1085 1085 A -> E (IN REF. 1).
FT CONFLICT 1123 1123 V -> C (IN REF. 1).
FT CONFLICT 1133 1133 L -> S (IN REF. 1).
FT CONFLICT 1144 1146 KSN -> NLI (IN REF. 1).
FT CONFLICT 1159 1168 RETKEQPKK -> TRKEQDKE (IN REF. 1).
FT CONFLICT 1179 1181 SKI -> ELKV (IN REF. 1).
FT CONFLICT 1184 1185 LE -> WK (IN REF. 1).
FT CONFLICT 1188 1204 LSQELSLNQYLKNRISG -> CHRKYLSILKQKNIR (IN REF. 1).
FT CONFLICT 1224 1224 P -> S (IN REF. 1).
FT CONFLICT 1228 1228 E -> Q (IN REF. 1).
FT CONFLICT 1253 1253 E -> O (IN REF. 1).
FT CONFLICT 1311 1323 PKESDINKLMLE -> LTKSLILTNGNAS (IN REF. 1).
FT CONFLICT 1400 1400 D -> H (IN REF. 1).
FT CONFLICT 1454 1554 SEQDLRLQKDLSETERKELLSSTIKQKQKQFNCMDLQ
NELRLREHIALKQAEEDVKNMASIEIKLTKQNKQKLLW
EREMERNDSDMQLQETLLE -> P (IN REF. 1).

FT CONFLICT 1568 1568 D -> V (IN REF. 1).
FT CONFLICT 1630 1646 DILKQLDHYTKVVEMLN -> SEARSLSYKSGNVD (IN REF. 1).
FT CONFLICT 1698 1704 MISSING (IN REF. 1).
FT CONFLICT 1725 1737 TLQLOMEQNSRNG -> NNTANGTRFKEW (IN REF. 1).
FT CONFLICT 1754 1757 FDDE -> LMM (IN REF. 1).
FT CONFLICT 1777 1777 D -> E (IN REF. 1).
FT CONFLICT 1788 1788 R -> T (IN REF. 1).
FT CONFLICT 1825 1825 S -> D (IN REF. 1).
FT CONFLICT 1882 1882 S -> W (IN REF. 1).
FT CONFLICT 1902 1904 FWK -> NSGRKLDADDL (IN REF. 1).
SQ SEQUENCE 1928 AA; 223634 MW; 6F54C7611F43DC9F CRC64;

alignment_scores:
Quality: 65.00 Length: 40
Ratio: 2.500 Gaps: 0
Percent Similarity: 65.000 Percent Identity: 40.000

alignment_block:

US-09-049-696-12 x MYS1_YEAST

Align seg 1/1 to: MYS1_YEAST from: 1 to: 1928

88 CTGATTGAATCAGTGATGGAACAGTTACCTGTGGAACACTACTGGATAA 137
|||||
526 LeuilegSerLysGlyProThrGlyValLeuProLeuLeuAspGI 542
138 TGGAGCAGGTGCTACTAGGAGCGGTGCTCAAGGATT 187
|||
542 uGluAlaValLeuProLysSerThrAspGluSerPheTyrSerLysLeuI 559

188 TCACAACTTATGACAGCAAT 207
: : : : :
559 leSerThrTrpAspGlnAsn 565

seq_name: SwissProt_39:FIG2_YEAST

seq_documentation_block:

ID FIG2_YEAST STANDARD; PRT; 1609 AA.
AC P25653;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE FACTOR INDUCED GENE 2.
GN FIG2 OR YCR089W OR YCR89W OR YCR1102.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92397594; PubMed=1523889;
RA Wilson C., Grisanti P., Frontali L.;
RT "The complete sequence of a 6146 bp fragment of Saccharomyces cerevisiae chromosome III contains two new open reading frames."
RL Yeast 8:569-575(1992).
CC -!- FUNCTION: REQUIRED FOR EFFICIENT MATING.
CC -!- INDUCTION: BY MATING PHEROMONES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X59720; CAA42254.1;
DR PIR; S19504; S19504.
DR PIR; S25345; S25345.
DR SGD; S0000685; FIG2.

SQ SEQUENCE 1609 AA; 166049 MW; 7D66AD7F85A7B852 CRC64;

alignment_scores:
Quality: 63.00 Length: 73
Ratio: 1.370 Gaps: 3
Percent Similarity: 63.014 Percent Identity: 30.137

alignment_block:

US-09-049-696-12 x FIG2_YEAST

Align seg 1/1 to: FIG2_YEAST from: 1 to: 1609

ACACGAAATTCACCCAGCCCTCTGCTAGTTATGCAAAATATTCGCCAAGG 53
|||||
682 ThrSerLysTyrProSerThrSerVal.....SerAsnProThrGlu.. 695
54 AGCTCCCAATTCACAGGCCAGTGCACAGCCCTGATTCGAATCAATCAGTGA 103
|||||
696 AlaserGlnHisValSerSerValAsnSerLeuThrAspPheThrS 712
104 ATGGAACACAGTTACCTTGGAACTACTGGATAATGGAGCAGGTGCTGAT 153
|||||
712 erAsnSerThrGluThrIleAlaValIleSerAsnIleHisLysThrSer 728
154 GCTACTAAGATGACGGTGTCTACTCAAGGTATTTCACAACTTATGACAC 203
: : : : :
729 SerAsnLysAsp.....TyrSerLeuThrThrThrGlnLeuLysTh 742
204 GAATGGTAGATACAGTGTA 222
: : : : :
742 rSerGlyLysGlnThrLeu 748

seq_name: SwissProt_39:LPXK_NEIMA

seq_documentation_block:

ID LPXK_NEIMA STANDARD; PRT; 344 AA.
AC Q99VE4;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TETRAACYLDISACCHARIDE 4'-KINASE (EC 2.7.1.130) (LIPID A 4'-KINASE).
GN LPXK OR NMA0872.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
KN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=72491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
FA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "complete DNA sequence of a serogroup A strain of Neisseria meningitidis 72491."
RL Nature 404:502-506(2000).
CC -!- FUNCTION: TRANSFERS THE GAMMA-PHOSPHATE OF ATP TO THE 4'-POSITION
CC OF A TETRAACYLDISACCHARIDE 1-PHOSPHATE INTERMEDIATE (TERMED DS-1-
CC P) TO FORM TETRAACYLDISACCHARIDE 1,4'-BIS-PHOSPHATE (LIPID IVA)
CC (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + 2,3-BIS(3-HYDROXYTETRADECANOYL)-D-
CC GLUCOSAMINYL-(BETA-D-1,6)-2,3-BIS(3-HYDROXYTETRADECANOYL)-D-
CC GLUCOSAMINYL BETA-PHOSPHATE -> ADP + 2,3,2',3'-TETRAKIS(3-
CC HYDROXYTETRADECANOYL)-D-GLUCOSAMINYL-1,6-BETA-D-GLUCOSAMINE 1,4'-
CC BISPHOSPHATE.
CC -!- PATHWAY: LIPID A BIOSYNTHESIS; SIXTH STEP.
CC -!- SIMILARITY: BELONGS TO THE LPXK FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -


```

[2]
SEQUENCE OF 1-114 FROM N.A.
MEDLINE=85160858; PubMed=3981640;
Christie G.E., Calendar R.;
"Bacteriophage P2 late promoters. II. Comparison of the four late
promoter sequences.";
J. Mol. Biol. 181:373-382(1985).
CC -1- FUNCTION: FORMS THE VIRUS CONTRACTILE TAIL SHEATH.
CC -1- SIMILARITY: 30% IDENTITY TO THE TAIL SHEATH PROTEIN OF PHAGE PS17,
CC AND BY EXTENSION, PROBABLY TO THE R-TYPE PYOCINS (ANTIBIOTICS)
CC PRODUCED BY P.AERUGINOSA STRAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AF063097; AAD03289.1; -
CC PIR: S26391; S26391.
CC INIT_MET 0
CC SEQUENCE 395 AA; 43012 MW; 83B6751D2B92168E CRC64:
SQ

```

```

alignment_scores:
  Quality: 62.00      Length: 79
  Ratio: 1.348        Gaps: 2
  Percent Similarity: 58.228      Percent Identity: 26.582

alignment_block:
  US-09-049-6996-12 x VPFL_BPP2      ..
  Align seg 1/1 to: VPFL_BPP2      from: 1 to: 395

```

1 GACACCAGCAAAATCCCCC...AGCCCTCTGGTAGTTTATGCAAATATT... 45
|||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
37 AspAlaGluThrPheProLeuAsnLysProValLeuIleThrAsnValG1 53
46CGCCAAGGAGCCTCCCCAATCTCACGCCGACAGTGTC A 82
::: ||| :::: ||| ||||||:::
53 nSerAlaIleSerLysAlaGlyLysLysGlyThrLeuAlaAlaSerLeuG 70
83 CAGCCCTGATTGAATCAGTCAGTAATGGAAAAACAGTTTACCTTGGAACTACG 132
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
70 InAlaIleAlaAspGlnSerLysProValThrValValMetArgValGlu 86
133 GATAATGGACGAGGTGCTGATGCTACTAGGATGACGGTGTCTACTCTACAAG 182
|||::: |||::: |||::: |||::: |||::: |||::: |||:::
87 AspGlyThrGlyAspaspGluGluThrLysLeuAlaGluThrValSerAS 103

183 GTATTTCACAACTTATGACACGAATGGTAGATACACT 219
::: |||::: |||::: |||::: |||::: |||::: |||:::
103 nIleIecGlyThrThrAspGluasnGlyGlnTrpYrThr 115

seq_name: SwissProt_39:ABC1_YEAST

seq_documentation_block:

ID	ABC1_YEAST	STANDARD	PRT: 501 AA.
AC	P27697;		
DT	01-AUG-1992 (Rel. 23, Created)		
DT	01-AUG-1992 (Rel. 23, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	DE ABC1 PROTEIN, MITOCHONDRIAL PRECURSOR.		

GN Saccharomyces cerevisiae (Baker's yeast).
 ABCL OR 1G113W.
 ON Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 [1]
 RN RN
 SEQUENCE FROM N.A.
 RP STRAIN-S9-12/50;
 RC

```

FX MEDLINE=91293073; PubMed=1648478;
RA Bousquet I., Dujardin G., Slonimski P.P.;
RT "AEC1, a novel yeast nuclear gene has a dual function in
RT mitochondria: it suppresses a cytochrome b mRNA translation defect
RT and is essential for the electron transfer in the bc 1 complex.";
RL EMBO J. 10:2023-2031(1991).
RN [2]
RN SEQUENCE FROM N.A.
RP Lauquin G.;
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN HAS A DUAL FUNCTION: IT SUPPRESSES A
CC CYTOCHROME B mRNA TRANSLATION DEFECT AND IS ESSENTIAL FOR THE
CC ELECTRON TRANSFER IN THE BC1 COMPLEX. IT IS POSSIBLE THAT ABC1
CC IS INVOLVED IN THE CORRECT FOLDING AND/OR ASSEMBLY OF CYTOCHROME
CC B AND THAT THIS FOLDING OCCURS DURING THE TRANSLATION PROCESS.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE ABC1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X59027; CAA41759.1; -
CC EMBL; Z72641; CAA96827.1; -
CC FIR; S16711; S16711.
CC SGD; S0003087; ABC1.
CC Mitochondrion; 1 ? MITOCHONDRION (POTENTIAL).
CC TRANSIT ? 501 ABC1 PROTEIN.
CC CHAIN ?
CC SEQUENCE 501 AA; 56741 MW; C91D1EE5A7ACC9C CRC64;
SQ

```

```

alignment_scores:
  Quality: 62.00      Length: 47
  Ratio: 2.138        Gaps: 1
  Percent Similarity: 61.702  Percent Identity: 36.170

alignment_block:
  US-09-049-696-12 x ABC1_YEAST      ..

  Align seg 1/1 to: ABC1_YEAST from: 1 to: 501

103  AATGGA AAAACAGTTACTTGGAACTACTGATAATGCAGAGTGCTGA 152
      |||||:||||| :|||:|||||:|||||:|||||:
353  AsnGlyAyrThrLysLysIleGluLeuLeuAspPheGlyAlaSerArgpr 371
      |||||:||||| :|||:|||||:|||||:|||||:
153  TGCTACTAAGGATGACGGTGCTCTACTCAAGGTATTTCCAACTTAT 198
      |||||:||||| :|||:|||||:|||||:|||||:
271  ophealaGluAspPheIleLeuLysTyraArgLysLeuLeuThrTyraLat 388

199  .....GACACGAATGTAGATACAGTCAGTAAAGTCCGG 231
      ||| :|||:||||| :|||:|||||:|||||:
388  hrLeuArgAspArgLysGlyAlaTyraGluMetSerValGln 401

```

seq name: SwissProt 39:CMF DICDI

```
seq_name: SWISSLOC_39:CMF_DICDI
seq_documentation_block:
```

ID	seq_documentation_BLOCK	STANDARD;	PRT;	571 AA.
CMF_DICDI				
P34090;				
AC				
DT	01-FEB-1994 (Rel. 28, Created)			
AT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	01-FEB-1994 (Rel. 28, Last annotation update)			
DT	01-JUN-1994 (Rel. 29, Last annotation update)			
DE	CONDITIONED MEDIUM FACTOR (CMF) (DENSITY-SENSING FACTOR).			
GN	CMFA.			
OS	Dictyostelium discoideum (Slime mold).			
OC	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.			
NCBI_TaxId=44689;				
11]				
RP	SEQUENCE FROM N. A. AND SEQUENCE OF 277-109. 319-331 AND			

```
RC STRAIN-AX4;
RX MEDLINE=92192452; PubMed=1547939;
RA Jain R., Yuen I.S., Taphouse C.R., Gomer R.H.;
RT "A density-sensing factor controls development in Dictyostelium.";
RL Genes Dev. 6:390-400(1992).
CC -1- FUNCTION: INVOLVED IN CELL DENSITY SENSING AND MIGHT SYNCHRONIZE
CC THE ONSET OF DEVELOPMENT BY TRIGGERING AGGREGATION WHEN A MAJORITY
CC OF THE CELLS IN A GIVEN AREA HAVE STARVED.
CC -1- PTM: N- AND O-GLYCOSYLATED.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z11691; CAA77749.1; -
DR PIR: S24482; S24482.
DR DICTYDB: DD02025; cmfa.
KW Glycoprotein.
FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 571 AA; 62510 MW; 9AB13293D0AF951C CRC64;

alignment_scores:
  Quality: 61.50 Length: 59
  Ratio: 1.538 Gaps: 2
Percent Similarity: 67.797 Percent Identity: 27.119

alignment_block:
US-09-049-696-12 x CMF_DICDI ..
  Align seg 1/1 to: CMF_DICDI from: 1 to: 571
58 TCCCAATTCAGGCCAGTGTACAGCC.....CTGATTGAATCACT 101
|||||
108 SerProileLeuLeuSerAspValSerAlaGluMetIleLeuGlyLeuPr 124
102 GAATGAAACAGTTACTTGGAACTACTGTAATGATGAGCAGCTGCTG 151
|||||
124 aspGlySerLysGluThrIleProMetPheAspGlyLeuHisAspA 141
152 ATGCTACTAAGGATGACGGTGTCTACTCAAGGTATTTCAACAACATGAC 201
|||||
141 spGluGlnAlaAsnAspGlyLeuPheGlyGlyTyIleAsnValSerGlu 157
202 ACGAATGGTAGATACAGTGTAAAGTG 228
|||||
158 Leu...GlyAsnHisAspLeuGlnVal 165

seq_name: SwissProt_39:EF1A_DICDI
seq_documentation_block:
ID EF1A_DICDI STANDARD; PRT; 456 AA.
AC P18624;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) (50 KDA ACTIN-BINDING PROTEIN)
DE (ABP-50).
DE GN
DE EF1A.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=AX3;
```

```

RX MEDLINE=91015340; PubMed=2215665;
RA Yang F., Demma M., Warren V., Dharmawardhane S., Condeelis J.;
RT "Identification of an actin-binding protein from Dictyostelium as
RT elongation factor la.";
RL Nature 347:494-496(1990).
CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS. IT IS ALSO AN ABUNDANT ACTIN FILAMENT BUNDLING
CC PROTEIN.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X55972; CAA39442.1; -
DR EMBL: X55973; CAA39443.1; -
DR PIR: S11665; S11665.
DR HSP: P07157; IAIIP.
DR DICTYDB: DD01027; efaa.
DR InterPro: IPR000795; GTP_EFTU.
DR Pfam: PF00009; GTP_EFTU; 1.
DR PRINTS: PR00315; ELONGATNFT.
DR PROSITE: PS00301; EFATOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding; Actin-binding.
FT NE_BIND 17 24 GTP (BY SIMILARITY).
FT NE_BIND 94 98 GTP (BY SIMILARITY).
FT NE_BIND 156 159 GTP (BY SIMILARITY).
SQ SEQUENCE 456 AA; 50066 MW; BCFF9E1B70C3D8A8 CRC64;

alignment_scores:
  Quality: 61.00 Length: 95
  Ratio: 1.196 Gaps: 4
Percent Similarity: 53.684 Percent Identity: 24.211

alignment_block:
US-09-049-696-12 x EF1A_DICDI ..
  Align seg 1/1 to: EF1A_DICDI from: 1 to: 456
1 GACACGACAAATCCCGAGCCCTCGTAGTTAT.....GCAAA 41
|||||
327 GluThrGluLysPheValAlaGlnValIleValLeuAsnHisProGlyG1 343
42 TATTCGCCAAGAGCGCTCCCAATTCTC.....A 70
|||||
343 nIleHisAlaGlyTySerProValLeuAspCysHisThrAlaHisIleA 360
71 GGGCCAGTGTCAAGCCCTGATTGAATGATGAGTGAATGGAATAAACA..... 114
360 lacLysPheThrGluIleValAspLysValAspArgThrGlyAla 376
115 .....GTTACCTTGGAACTACTGATATGGAGCAGG 146
377 ValValAlaLysGluGlyThrAlaAlaValLeuLysAsnGlyAspAl 393
147 TGCT.....GATGCTACTAGGAGCGGTGCTCTACTCAAGGTATTTC 190
|||||
393 AlaMetValGluLeuThrProSerArgProMetCysValGluSerPheT 410
191 CAACCTATGACACGAATGGTAGATACAGTGTAATAA 225
410 hrGluTyProProLeuGlyArgPheAlaValArg 421

seq_name: SwissProt_39:YP58_YEAST
```

```
seq_documentation_block:
ID YP58_YEAST STANDARD; PRT; 758 AA.
AC G99299;
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE HYPOPHETICAL 84.8 KDA PROTEIN IN CDC60-PRM4 INTERGENIC REGION.
GN YPL158C OR P2570.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=288C / AB972;
RA MEDLINE=97103777; PubMed=8948103;
RX Purnelle B., Coster F., Goffeau A.;
RT "The sequence of 55 kb on the left arm of yeast chromosome XVI
RT identifies a small nuclear RNA, a new putative protein kinase and two
RT new putative regulators.";
RL Yeast 12:1483-1492(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X96770; CAA65563.1; -
DR EMBL; 273514; CAA97863.1; -
DR SGD; S0006079; YPL158C.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 253 263 POLY-SER.
FT DOMAIN 648 711 COILED COIL (POTENTIAL).
FT DOMAIN 652 658 POLY-GLU.
FT DOMAIN 664 669 POLY-GLU.
FT DOMAIN 686 690 POLY-GLU.
SQ SEQUENCE 758 AA; 84845 MW; 3B4FA92B91C87F2B CRC64;

alignment_scores:
Quality: 61.00 Length: 63
Ratio: 1.452 Gaps: 4
Percent Similarity: 66.67 Percent Identity: 36.508

alignment_block:
US-09-049-696-12 x YP58_YEAST ..
Align seg 1/1 to: YP58_YEAST from: 1 to: 758
19 AGCCCTCTGGTGGTATAT.....GCAATAT 44
|||||
407 SerProLeuValSerTyrThrProSerLeuArgArgThrAsnSerSerI 423
45 TCGCCAAGGACCTCCCAATTCACGGCCAGTGTACACGCCCTGATTG 94
|||||
423 eArgAlaAlaSer...lleLeuThrAlaSerAlaThr.....Met 437
95 AATCAGTGAATGGAAACAGCTTACCTTGGAACTGGAATGAGGACGA 144
|||||
437 hrProAlaAsnAsnLysAsnSerPheIleSerValProAspAsnValSer 453
145 GGTGCTGATGCTACTAAGGATGACGGTGTCTACTCAAGG 183
|||||
454 HisAla...ValThrArgAsnSerSerMetTyrSerArg 465

seq_name: SwissProt_39:CBPA_CLOCL

seq_documentation_block:
ID CBPA_CLOCL STANDARD; PRT; 1848 AA.
AC P38058;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CELLULOSE BINDING PROTEIN A PRECURSOR.
GN CBPA.
OS Clostridium cellulovorans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1493;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=92228810; PubMed=1565642;
RX Shoseyov O., Takagi M., Goldstein M.A., Doi R.H.;
RT "Primary sequence analysis of Clostridium cellulovorans cellulose
RT binding protein A.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3483-3487(1992).
CC -!- FUNCTION: BINDS TO CELLULOSE FIBERS AND COORDINATES CELLULOSE
CC ENZYMES.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- PTM: CONTAINS ABOUT 10% OF CARBOHYDRATES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M73817; AAA23218.1; -
DR PIR; A44140; A44140.
DR HSSP; Q06851; INBC.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR002102; Cohesin.
DR Pfam; PF00942; CBD_3; 1.
DR Pfam; PF00963; Cohesin; 9.
DR ProDom; PD001947; CBD_3; 1.
KW Cellulose degradation; Cell wall; Repeat; Signal; Glycoprotein.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 1848 CELLULOSE BINDING PROTEIN A.
FT DOMAIN 29 189 CELLULOSE-BINDING (BY SIMILARITY).
SQ SEQUENCE 1848 AA; 189152 MW; 85FA6CE6F771AF1A CRC64;

alignment_scores:
Quality: 61.00 Length: 56
Ratio: 1.694 Gaps: 1
Percent Similarity: 64.286 Percent Identity: 26.786

alignment_block:
US-09-049-696-12 x CBPA_CLOCL ..
Align seg 1/1 to: CBPA_CLOCL from: 1 to: 1848
13 TTCGCCAGCCCTCTGGTAGTTTATGCAATATTCGCCAAGGAGGCTCCCC 62
|||||
1748 PheAspSerSerValLeuThrTyrValGlyThrAlaGlyThrSer.. 1763
63 AATTCTCAGGCCAGTGTACACGCCCTGATTGAATCAGTGAATGAAAAA 112
|||||
1764 ....lleLysAsnProAlaValAsnPheSerSerGlnLeuAsnGlyAsnT 1779
113 CAGTTACTTGGAACTACTGGAATAATGAGCAGGTGCTGCTACTAAG 162
|||||
1779 hrIleThrLeuLeuPheAspAsnThrIleGlyAsnGluLeuIleThr 1795
163 GATCAGGTGCTACTCA 180
|||||
1796 AlaAspGlyGlnPheAla 1801

seq_name: SwissProt_39:VGL_BP1F1

seq_documentation_block:
```

```
ID VGI_BP1F1 STANDARD; PRT; 353 AA.
AC O80299;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MATURATION PROTEIN.
GN I OR 1.
OS Bacteriophage If1.
OC Viruses; ssDNA viruses; Inoviridae; Inovirus.
OX NCBI_TaxID=10868;
RN [1]
RP SEQUENCE FROM N.A.
RA Hill D.F., Hughes G., McNaughton J.C., Stockwell P.A., Petersen G.B.;
RT "DNA sequence of the filamentous coliphage If1.";
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE EXACT FUNCTION OF THIS PROTEIN IS NOT KNOWN ALTHOUGH
CC IT MAY BE INVOLVED IN PHAGE ASSEMBLY (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U02303; AAC62157.1; -.
KW Phage maturation; ATP-binding.
FT NP_BIND 8 15 ATP (POTENTIAL).
SQ SEQUENCE 353 AA; 39990 MW; BB8CB367CCFA949E CRC64;
```

alignment_scores:
Quality: 60.50 Length: 73
Ratio: 1.476 Gaps: 2
Percent Similarity: 56.164 Percent Identity: 30.137

alignment_block:
US-09-049-696-12 x VGI_BP1F1 ..
Align seg 1/1 to: VGI_BP1F1 from: 1 to: 353
7 AGCAATTCCTCCAGCCCTCTGGTAGTTATGCAATATTGCGCCAGGAGC 56
|||||::||| ||| ||| :::: |||
167 SerLysLeuProLeuProLysValHisValGlyIleValLysTyrGlyAs 183
57 CTCCCAATTCTCAGGCGCCAGCTGCACGCCCTGATTTGAATCAGTGAATG 106
||||| |||::|||::: :::
183 pSerPro.....GlnSerMetThrValGluArgTyrThrG 197
107 GAAAAACAGTTACCTTGGAACTACTCGATATGAGACAGGTGCTGATGCT 156
||::: :::: ||| ::|||
197 LysArgAspLeuTyrAlaAlaTyrAspThrLysGlnAlaPheSerAspAla 213
157 ACTAAGGATGACGGTCTCTACTCAAGGTATTTTCACAACTTATGACACGAA 206
::: ::::: |||::||| ||| :::::
214 TyrGluHisSerSerPhe.....SerTyrLeuThrProTyrLeuSerHI 228
207 TGCTAGATACAGTGTAAAA 225
:|||||::|||
228 sGlyArgTyrAlaValLys 234

Score	Strd	Orig	Zscore	Escore	Len	Documentation
p1r2:S702068	+	339.00	706.95	6.6e-33	913	Gub-ECM-1 protein - mouse
p1r2:TJ02005	+	243.00	507.31	3.2e-21	913	Luo-ECAM-1 protein - bovine
p1r2:S44257	+	79.50	152.79	0.2473	651	phosphotransferase system enzyme
p1r2:TJ702010	+	69.00	132.81	4.64	450	conserved hypothetical protein
p1r2:TJ21287	+	69.00	131.85	4.65	507	hypothetical protein P23B12.5
p1r2:TJ84292	+	69.00	130.43	4.68	606	signal-transducing histidine kinase
p1r2:S72966	-	68.50	129.59	5.37	587	probable recombination protein
p1r2:F65207	+	67.50	129.27	7.07	465	hydh protein - Escherichia coli
p1r2:TJ30538	+	66.00	114.94	11.18	1847	resistance protein RGC2J - gal
p1r2:S46773	+	65.00	112.41	14.81	1928	myosin heavy chain - yeast
p1r2:TJ70640	+	64.50	126.93	16.14	275	hypothetical protein RV0687 - M
p1r2:JC1370	+	64.00	135.93	17.93	78	alpha-amylase inhibitor AI-409 -
p1r1:S75524	+	63.50	122.19	21.55	378	sensory transduction system reg
p1r2:TJ31179	+	63.50	119.84	21.73	507	hypothetical protein 480 - Shig
p1r2:F82198	+	63.50	117.02	21.94	721	probable toxin secretion transp
p1r2:S49387	+	63.00	121.69	24.74	351	replication-associated protein
p1r2:B24336	+	63.00	121.69	24.74	351	replication-associated protein
p1r2:S25345	+	63.00	109.49	25.81	1609	probable membrane protein YCRQ
p1r2:F81170	+	62.50	120.76	28.44	344	tetraacyldisaccharide 4'-kinase
p1r2:C81933	+	62.50	120.76	28.44	344	probable tetraacyldisaccharide
p1r2:S26391	+	62.00	118.54	32.85	396	tail sheath protein - phage P2
p1r2:TJ6711	+	62.00	116.66	33.06	501	ABG1 protein precursor - yeast
p1r1:S75536	+	62.00	116.16	33.12	533	NADH dehydrogenase (ubiquinone)
p1r2:TJ7136	+	62.00	116.09	33.13	538	hypothetical protein Y54E5A.4
p1r2:TJ0637	-	62.00	105.05	34.43	2133	hypothetical protein MC035R -
p1r2:C72539	-	61.50	126.08	36.67	135	hypothetical protein APEI1600 -
p1r2:S83729	+	61.50	115.41	38.06	511	phosphoribosylaminoimidazolecar
p1r2:A83741	+	61.50	114.52	38.17	571	conditioned medium factor - sli
p1r2:A42138	+	61.50	114.52	38.17	571	conditioned medium factor - sli
p1r2:TJ30856	+	61.50	109.72	38.82	1039	protein F2 - Streptococcus pyo
p1r2:S14993	+	61.00	125.29	42.14	130	extensin class I (clone ABC-1) p
p1r2:E81179	+	61.00	118.72	43.11	295	sermidine/putrescine ABC trans
p1r2:A81927	+	61.00	118.72	43.11	295	probable polyamine permease en
p1r2:S11665	+	61.00	115.23	43.64	456	translation elongation factor e
p1r2:S65169	+	61.00	111.16	44.26	758	hypothetical protein vPL158c -
p1r2:TJ02850	+	61.00	105.76	45.10	1487	hypothetical protein L1439.2
p1r2:A44811	+	61.00	105.59	45.13	1518	glucosyltransferase (EC 2.4.1.1
p1r2:A44140	+	61.00	104.91	45.38	1848	cellulose-binding protein A pr
p1r2:S44424	+	60.50	119.73	49.23	227	phosphoribosylaminoimidazolec
p1r2:TJ75475	+	60.50	115.23	50.01	398	glucose-fructose oxidoreductase
p1r2:S62528	+	60.50	114.59	50.12	431	hypothetical protein SPCC584.14

A;Molecule type: mRNA

A; RESIDUES: 1-903 <ELB>
A; Cross-references: EMBL:AF001261: NID:Q2623762: PTD:AAB86529.1: PTD:Q2623763

A:Residues: 1-793,'SGSFP','M',862,'RF',865-866,'Q',868,'AKVLEIQ',876,'QHQ',880,'FQ' <EL2>
A:Cross-references: EMBL:AF001263; NID:g2623766; PIDN:AAB86531.1; PID:g2623767
A:Experimental source: lung
A:Accession: T02171
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-792,'ES' <EL3>
A:Cross-references: EMBL:AF001262; NID:g2623764; PIDN:AAB86530.1; PID:g2623765
A:Experimental source: lung

alignment_scores:
Quality: 243.00 Length: 77
Ratio: 3.919 Gaps: 0
Percent Similarity: 80.519 Percent Identity: 59.740

alignment_block:
US-09-049-696-12 x T02205 ..

Align seg 1/1 to: T02205 from: 1 to: 905

```
4 ACCAGCAAAATCCCGCCCTCTGGTAGTTATGCAAAATATTCGCCAAGG 53
|||||  ::::|  |||||  |||||  |||||  |||||  |||||  |||||
609 ThrAlaHisTyrProSerPrometIleValTyrAlaGlnValSerGlnG 625
1  |||||  |||||  |||||  |||||  |||||  |||||  |||||
625 yPheLeuProValLeuGlyIleSerValIleAlaIleIleGluThrGluA 642
104 ATGGAANAACAGTACTCTCGAAGCTACTGATGATGATGATGATGATG 153
|||||  |||||  |||||  |||||  |||||  |||||  |||||
642 spGlyHisGlnValThrLeuGluLeuTyrAspAsnGlyAlaGlyArgAsp 658
154 GCCTACTAAGGATGACGGTGCTCTACTCAAGGTATTTACACACTTATGACAC 203
|||||  |||||  |||||  |||||  |||||  |||||  |||||
659 ThrValIleAsnAspGlyIleTyrSerArgTyrPheThrAspTyrTyrG 675
204 GAATGGTAGATACAGTCTAAAGTGGCGGCT 234
|||||  |||||  |||||  |||||  |||||  |||||  |||||
675 yAsnGlyArgTyrSerLeuIleValHisAla 685
```

seq_name: pir2:S44257

seq_documentation_block:
C:Species: Pedicoccus pentosaceus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 16-Jul-1999
C:Accession: S44257
R:Leenhouts, K.K.J.; Bolhuis, A.A.; Kok, J.J.; Venema, G.G.
submitted to the EMBL Data Library, April 1994
A:Description: The sucrose and raffinose operons of *Pedococcus pentosaceus* PP1.0.
A:Reference number: S44252
A:Accession: S44257
A:Molecule type: DNA
A:Residues: 1-651 <LEE>
A:Cross-references: EMBL:Z32771; NID:g493728; PIDN:CAA83668.1; PID:g475968
C:Genetics:

A:Gene: scrA
C:Superfamily: phosphotransferase system enzyme II sucrose-specific; phosphotransferase
C:Keywords: phosphotransferase
F:488-640/Domain: phosphotransferase system glucose-specific enzyme II, factor III homol

alignment_scores:
Quality: 79.50 Length: 87
Ratio: 1.622 Gaps: 4
Percent Similarity: 56.322 Percent Identity: 29.885

alignment_block:
US-09-049-696-12 x S44257 ..

Align seg 1/1 to: S44257 from: 1 to: 651

```
1 GACACCAGCAAAATCCCGCCCTCTGGTAGTTATGCAAAATATTCGCCCA 50
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
475 AspGlnValIleSerProAlaProThrSerThrValIleAsnValAsn.. 490
51 AGGAGCCTCCCAATTCCTCAGGCCAGTGTCCACAGCCCTGATTGAATCA. 99
:::  ||:::  |||  |||||  |||  |||||  |||  |||||
491 .....AspGluIleIleSerAlaProValThrGlyAlaSerGluSerL 505
|||||  ::::|  |||||  |||||  |||||  |||||  |||||
100 .....GTGAATGGAAAAACAGTTACCTTGGAACTACTGGATAATGGA 141
|||||  ::::|  |||||  |||||  |||||  |||||  |||||
505 euLysGlnValAsnAspGlnValPheSerAlaGluIleMetGlyLysGly 521
142 GCAGGTGCTGATGCTACTTAAGGAT.....GACGGTGT 173
|||||  ::::|  |||||  |||||  |||||  |||||  |||||
522 AlaAlaIleValProSerSerAspGlnValValAlaProAlaAspGlyVa 538
174 CTACTCAAGGTATTTCACAACTTATGACACGAATGGTAGATACAGTGA 223
|||  :::  |||||  |||||  |||||  |||||  |||||
538 IleThr.....ValThrTyrAspSerHisAlaTyrGlyIleL 552
224 AAGTCGGGGCT 234
|||  |||
552 yThrThrAla 555
```

seq_name: pir2:G70210

seq_documentation_block:
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: G70210
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*.
A:Reference number: A70100; MUID:98065943
A:Accession: G70210
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-450 <KLE>
A:Cross-references: GB:AF000790; NID:g2690224; PIDN:AAC66261.1; PID:g2690260; TIGR:BB
A:Experimental source: strain B31
C:Genetics:
A:Genome: plasmid

alignment_scores:
Quality: 69.00 Length: 54
Ratio: 2.226 Gaps: 2
Percent Similarity: 57.407 Percent Identity: 35.185

alignment_block:
US-09-049-696-12 x G70210 ..

Align seg 1/1 to: G70210 from: 1 to: 450

```
34 TATGCAAAATTCGCCAAGGAGCTCCCAAT.....CTCAG 71
|||||  |||  |||  |||  |||  |||  |||  |||
308 TyrAlaTyrIleTyrGlnAspGlnIleProValSerAspSerLeuMetLe 324
72 GGCCAGTGTACAGCCCTGATTCAGTGAATGGAATGGAACACAGTTACCT 121
|||||  ::::|  |||||  |||||  |||||  |||||  |||||
324 uAlaSerIleGlnValIleGluAsnPheAsnValAsnThrValTyrI 341
122 TGGAACTACTGGATAATGAGCAGGTGCTGCTACTACTAAGGATGACGGT 171
::|  |||  |||||  |||||  |||||  |||||  |||||
341 leGlu.....GluArgAspSerThrLysGlyAspGly 351
172 GTCTACTCAGG 183
:::  ::::|
352 IleLeuThrLys 355
```

seq_name: pir2:T21287

seq_documentation_block:
 hypothetical protein F23B12.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
 C:Accession: T21287
 R:Wild, A.
 submitted to the EMBL Data Library, July 1996
 A:Reference number: Z19402
 A:Accession: T21287
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-507 <WIL>
 A:Cross-references: EMBL:Z77659; PIDN:CAB01163.1; GSPDB:GN00023; CESP:F23B12.5
 A:Experimental source: clone F23B12
 C:Genetics:
 A:Gene: CESP:F23B12.5
 A:Map position: 5
 A:Introns: 72/71; 314/2; 465/2
 C:Superfamily: dihydrolipoamide acetyltransferase; lipoyl/biotin-binding homology

alignment_scores:
 Quality: 69.00 Length: 74
 Ratio: 1.533 Gaps: 3
 Percent Similarity: 60.811 Percent Identity: 32.432

alignment_block:
 US-09-049-696-12 x T21287 ..

Align seg 1/1 to: T21287 from: 1 to: 507

```

7  AGCAAAATCCCGAGCCCTCTGTTATGCAAAATATTCGCCAAGGAGC 56
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
2  SerLysPheProValProLeuArgThrIleGlyGly..... 13

57  CTCGCCAATTCACAGGCCAGTCGACAGCCCTGATTCAGTCAAGTGAAT. 105
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
14  .....LeuArgProSerThrThrAlaAlaIleSerAlaAlaAsnI 27

106  ..GGAAACACAGTACCTTGGAACTACTGGATAATGGACAGGTGCTGAT 153
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
27  leGlyPheThrGlnSerSerArgAlaLeuSerThrGlyAla..... 40

154  GCTACTAAGTAGGACGGTCTCTACTCAAGGTATTCACAACTTATGACAC 203
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
41  AlaAlaLysSerSerGlyLeuValGlyGlnValAlaArgGlnTyrProAs 57

204  GAATGGTAGATACACTGTAAAA 225
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
57  nAlaAlaAlaPheSerIleLys 64

```

seq_name: pir2:A84292

seq_documentation_block:
 signal-transducing histidine kinase homolog [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: A84292
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
 Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li
 A:Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: A84160; MUID:20504483
 A:Accession: A84292
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-606 <STO>
 A:Cross-references: GB:AE004437; NID:g10580883; PIDN:AAG19701.1; GSPDB:GN00138

C:Genetics:
 A:Gene::kinA1

alignment_scores:
 Quality: 69.00 Length: 94
 Ratio: 1.468 Gaps: 3
 Percent Similarity: 50.000 Percent Identity: 23.404

alignment_block:
 US-09-049-696-12 x A84292 ..

Align seg 1/1 to: A84292 from: 1 to: 606

```

7  AGCAAAATCCCGAGCCCTCTGTTATGCAAAATATTCGCCAAGGAGC 30
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
478  SerSerPheProSerAlaAlaValThrValAlaSerGlySerIleProAspVa 494

31  .....GTTTATGCAAAATATTCGCC 49
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
494  lThrValHisGlyAsnAsnMetLeuAspSerValPheArgAsnLeuLeuL 511

50  AAGGAGCC.....TCCCAATTCAGGCCAGT 78
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
511  ysAsnAlaIleGlnHisAsnAspGlnAspSerPro.....Asp 523

79  GTCACAGCCCTGATGATGATGATGATGATGATGATGATGATGATGAT 128
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
524  ValThrValSerAlaAspCysThrAspAspAsnThrAlaThrIleHisVa 540

129  ACTGGATAATGGACAGGTGCTGATGCTACTAAGGATGACGGTGTCTACT 178
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
540  lAlaAspAsnGlyProGlyValProAspAlaGlnLysAspAlaIlePheG 557

179  CAAGGTATTCACAACTATGACACGAATGCT 210
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
557  lyLysGlyGluThrGlyLeuAspSerAspGly 567

```

seq_name: pir2:S72966

seq_documentation_block:
 probable recombination protein recN - Mycobacterium leprae
 C:Species: Mycobacterium leprae
 C>Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
 C:Accession: S72966
 R:Smith, D.R.; Robison, K.
 submitted to the EMBL Data Library, November 1993
 A:Description: Mycobacterium leprae cosmid L247.
 A:Reference number: S72589
 A:Accession: S72966
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-587 <SMI>
 A:Cross-references: EMBL:U00021; NID:g467141; PIDN:AAA50914.1; PID:g467150
 C:Genetics:
 A:Gene: recN
 C:Superfamily: recN protein

alignment_scores:
 Quality: 68.50 Length: 46
 Ratio: 1.803 Gaps: 2
 Percent Similarity: 82.609 Percent Identity: 39.130

alignment_block:
 US-09-049-696-12/rev x S72966 ..

Align seg 1/1 to: S72966 from: 1 to: 587

```

158  GTAGCATCAGCACCTCTCCATATTCAGTAGTTCACAGGTAACCTGTTT 109
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
259  MetGlyGlnAlaLysAlaAlaLeuGluSerThrAspAlaThrLeuAr 275

```

[illegible]

C:Accession: F65207; A33862
R:Platner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: F65207
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-465 <BLAST>
A:Cross-references: GB:AE000473; GB:U00096; NID:q2367336; PIDN:AAC76977.1; PID:gi1790436,
A:Experimental source: strain K-12, substrain MGL655
R:Stoker, K.; Reijnders, W.N.M.; Oltman, L.F.; Stouthamer, A.H.
J. Bacteriol. 171, 4448-4456, 1989
A:Title: Initial cloning and sequencing of *hydHG*, an operon homologous to *ntfBC* and regu
A:Reference number: A33862; MUID:89327164
A:Accession: A33862
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 328-360, 'IV', 362-382, 'TKA', 386,388-465 <STO>
A:Cross-references: GB:M28369; NID:gi146427; PIDN:AAA24003.1; PID:gi146427

C;Genetics: ..
A;Gene: hydH ..
C;Keywords: transmembrane protein ..

alignment_scores: ..
Quality: 67.50 Length: 57
Ratio: 2.045 Gaps: 1
Percent Similarity: 57.895 Percent Identity: 35.088

alignment_block: ..
US-09-049-696-12 x F65207 ..

```

25  CTGGTAGTTTATGCAAAATATTCCCAAGAGAGCCTCCCAATTCTCAGGCG 74
    ||| :||| ||| |||::: |||
362  LeuAsnLeuTyLeuAsnAlaIleGlnAlaIleGlyGlnHisGlyValIl 378
    ||| :||| ||| |||::: |||
75  CAGTGCACAGCCCTGATTGAATCAGTGAATGGAAGAAACAGTTACCTTGG 124
    ||||| ||||| ||| :||| |||
378  eSerValThrAla.....SerGluSerGlyAlaGlyValLysIleS 392
    ||||| ||||| ||| :||| |||
125  AACTACTGGTAAATGAGCAGGCTGCTACTACTAGGATGACGGTGTC 174
    :|||: |||::||| ||| ||| :|||: |||::|||
392  erValThrAspSerGlyLysGlyIleAlaAlaAspGlnLeuAspAlaIle 408
    :|||: |||::||| ||| ||| :|||: |||::|||
175  TACTCAAGGTATTTCACAAC 195
    :|||: |||::||| ||| ||| :|||: |||::|||
409  PheThrProTyrrPheThrThr 415
    :|||: |||::||| ||| ||| :|||: |||::|||

seq_name: pir:T30558

seq_documentation_block:
resistance protein RGC2V - garden lettuce
C:Species: lactuca sativa (garden lettuce)
Oct-1999 #text change 20-Oct-1999

```

•

n PC

ma

Yes

DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* s.

A:Reference number: S74322; MUID:97061201

A:Accession: S75524

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-378 <KAN>

A:Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BAA18085.1; PID:g165316

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: sensory transduction system regulatory protein; response regulator homolog

C:Keywords: phosphoprotein; transcription regulation

F:8-119/Domain: response regulator homology <RRH>

F:56/Binding site: phosphate (Asp) (covalent) #status predicted

alignment_scores:
Quality: 63.50 Length: 37
Ratio: 2.352 Gaps: 1
Percent Similarity: 72.973 Percent Identity: 35.135

alignment_block:
US-09-049-696-12 x S75524 ..

Align seg 1/1 to: S75524 from: 1 to: 378

76 AGTCACAGCCCTGATTGAATCAGTGAATGGAAGAAACACAGTACCTTGGG 125
|||||

290 SerlethrlethrValThrAlaIleAsp...GlnThrleuthrCysAr 305
.....

126 ACTACTGGTAATGAGCAGCGTGTGCTACTAAGGATGCGGTGTCT 175
.....

305 gVallyAspAsnGlyValGlyValAspLeuAlaIleAlaAspArgValp 322
.....

176 ACTCAAGGTAT 186
.....

322 heGluArgTyr 325
.....

seq_name: pir2:T31179

seq_documentation_block:

hypothetical protein 480 - *Sphingomonas aromaticivorans* plasmid pNL1

C:Species: *Sphingomonas aromaticivorans*

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000

C:Accession: T31179

R:Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Sensen, C.W.; G

submitted to the EMBL Data Library, July 1998

A:Description: Complete sequence of a 184 kb catabolic plasmid from *Sphingomonas aromati*

A:Reference number: Z20992

A:Accession: T31179

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-507 <ROM>

A:Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378320; PIDN:AAD03903.1

C:Genetics:

A:Genome: plasmid pNL1

A>Note: orf480

alignment_scores:
Quality: 63.50 Length: 57
Ratio: 1.924 Gaps: 1
Percent Similarity: 57.895 Percent Identity: 28.070

alignment_block:
US-09-049-696-12 x T31179 ..

Align seg 1/1 to: T31179 from: 1 to: 507

4 ACACGAAATCCCGAGCCCTGCTAGTTATGCAAAATATCGCCAGG 53
|||||

26 ThrAlaMetTyrValAspProLeuThrIleTyrArgGluTyrValGlnAs 42
.....

54 AGCCTCCCCA.....ATTCTCAGGCCCA 76
|||||

42 nAlaAlaAlaIleAspAlaArgAlaAlaGlyLeuLeuArgAla 59
.....

77 GTGTCACAGCCCTGATTGAATCAGTGAATGGAAGAAACACAGTACCTTGGAA 126
|||||

59 spGluGlyAlaIleAlaIleThrIleAspProThrThrArgThrIleThr 75
|||||

127 CTACTGGAATGAGCAGGT 147
.....

76 IleArgAspAsnGlyValGly 82
.....

seq_name: pir2:F82198

seq_documentation_block:

probable toxin secretion transporter VCL146 [imported] - *Vibrio cholerae* (strain N169

C:Species: *Vibrio cholerae*

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: F82198

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A:Reference number: A82035; MUID:20406833

A:Accession: F82198

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-721 <HEI>

A:Cross-references: GB:AE004223; GB:AE003852; NID:g9655942; PIDN:AAF94603.1; GSPDB:GN

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VCL146

A:Map position: 1

C:Superfamily: hemolysin secretion protein B; ATP-binding cassette homology

alignment_scores:
Quality: 63.50 Length: 88
Ratio: 1.323 Gaps: 3
Percent Similarity: 54.545 Percent Identity: 27.273

alignment_block:
US-09-049-696-12 x F82198 ..

Align seg 1/1 to: F82198 from: 1 to: 721

28 GTAGTTTATGCAAAATATT.....CGCCAAGGAGCCTCCCC 62
|||||

53 ValSerTyrAlaAsnLeuThrAspMetGluMetArgGluAlaAlaAspTyr 69
.....

63 AATT...CTCAGGGCAGTGCACA..... 84
.....

69 rLeuGlyLeuLysSerGlnIleThrLysLeuSerIleGlnAlaPheAsnT 86
.....

85GCCCTGATTGATCAGTGAATGGAAGAAACACAGTTACC 120
|||||

86 hrLeuProLeuProAlaLeuIleGluTyrGlnGlyTrpLysValMet 102
.....

121 TTGGAACACTGTAATGAGCAGGTGCTGATGCTACTAAGGATGACGG 170
.....

103 IleGlnThrAspAspAsnGlyTrpThrAlaTyrAspProAlaThrAsp 119
.....

171 TGTCTACTCAAGGTATTTCACACTTATGACACGAAATGCTAGATACAGTG 220
.....

119 rLeuHisThrLeuSerPheThrSerAlaGluSerThrAlaHisTyrLysV 136
.....

221 TAAAAGTGGGGCT 234
.....

136 alMetLeuIleAla 140
.....


```
; APPLICANT: Rosen, Craig
; TITLE OF INVENTION: Colon Specific Genes and Proteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; ADDRESSEE: Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07289
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-265
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-07289-9

alignment_scores:
  Quality: 111.00      Length: 21
  Ratio: 5.286        Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-049-696-12 x PCT-US95-07289-9
..
Align seg 1/1 to: PCT-US95-07289-9 from: 1 to: 228

172 GTCTACTCAAGGTATTTCACAACTTATGACACGAATGGTAGATACAGTGT 221
|||||
1 ValTyrSerArgTyrPhetThrThrTyrAspThrAsnGlyArgTyrSerVa 17

222 AAAAGTCGGCGCT 234
|||||
17 LlysValArgAla 21

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-597-236-2

seq_documentation_block:
; Sequence 2, Application US/08597236
; Patent No. 5733765
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/597,236
; FILING DATE:
; CLASSIFICATION: 426
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 484 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-597-236-2

alignment_scores:
  Quality: 63.50      Length: 76
  Ratio: 1.411        Gaps: 3
  Percent Similarity: 59.211      Percent Identity: 28.947

alignment_block:
US-09-049-696-12 x US-08-597-236-2
..
Align seg 1/1 to: US-08-597-236-2 from: 1 to: 484

34 TATGCAATATTGCCAAGAGAGCTCCCA.....ATTCTCAGGCCAG 77
|||
174 TyrAspAsnLeuLysSerGlyLysSerLysAlaMetValLeuSerGlyse 190

78 TGTCAAGCCCTGATTGATCATCATG.....AATGGAA 109
|
190 rTyAlaSerLeuLeuSerValAspSerAsnTyrAlaSerAsnLeu 207
|||||
110 AAACAGTT...ACCTTGGAACTACTGGATAATGGACAGGTGCTGATGCT 156
|||||
207 yThrIleTyrThrTyrLysIleLysLysLysAsnSerAsnSerAlaAsn 223

157 ACTAAGGATGACGGTGTCTACTCAAGGTATTTCACAACTTATGACACGAA 206
|||||
224 GlnValAspSerArgValPheAsnIleTyrIleSerGlyIleAspThrTy 240

207 TGGTAGATACAGTGTAAAGTCGGGCT 234
|||
240 rGlyProIleSerThrValSerArgSer 249

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-746-682A-2

seq_documentation_block:
; Sequence 2, Application US/08746682A
; Patent No. 5786184
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```



```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,682A
; FILING DATE: 14-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/597,236
; FILING DATE: 20-JUN-1995
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fanucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 484 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-746-682A-2

```

```

alignment_scores:
  Quality: 63.50      Length: 76
  Ratio: 1.411       Gaps: 3
  Percent Similarity: 59.211   Percent Identity: 28.947

```

```

alignment_block:
US-09-049-696-12 x US-08-746-682A-2 ..

```

Align seg 1/1 to: US-08-746-682A-2 from: 1 to: 484

```

34  TATGCAATATTGCGCAAGAGCCTCCCA.....ATTCTCAGGCGCAG 77
   ||| |||:|||||:||||| ||| |||:|||||:|||||
174 TyrAspAsnLeuLysSerGlyLysSerLysAlaMetValLeuSerGlySe 190
   ||| |||:|||||:||||| ||| |||:|||||:|||||
78  TGTCAAGCCCTGATGAATCAGTG.....AATGAA 109
   ||| |||:|||||:||||| ||| |||:|||||:|||||
190 rTyAlaSerLeuLeuGluSerValAspSerAsnTyrAlaSerAsnLeuL 207
   ||| |||:|||||:||||| ||| |||:|||||:|||||
110 AAACAGTT...ACCTGGAACCTACTGGATAATGGACGAGGTGCTGATGCT 156
   ||| |||:|||||:||||| ||| |||:|||||:|||||
207 ySThrIleTyrThrTyrLysIleLysLysLysAsnSerAsnSerAlaAsn 223
   ||| |||:|||||:||||| ||| |||:|||||:|||||
157 ACTAAGGATGACGGTGCTACTCAGGTATTTCACAACTTATGACACGAA 206
   ||| |||:|||||:||||| ||| |||:|||||:|||||
224 GlnValaspSerArgValPheAsnIleTyrIleSerGlyIleAspThrTy 240
   ||| |||:|||||:||||| ||| |||:|||||:|||||
207 TGGTAGATACAGTGTAAGTGCAGGCT 234
   ||| |||:|||||:||||| ||| |||:|||||:|||||
240 rGlyProIleSerThrValSerArgSer 249

```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-779-113-2

```

seq_documentation_block:
; Sequence 2, Application US/08779113
; Patent No. 5948891
; GENERAL INFORMATION:
; APPLICANT: Staunton, Donald E.
; APPLICANT: Harris, Edith S.
; TITLE OF INVENTION: Cytoplasmic Modulators of Integrin
; TITLE OF INVENTION: Binding
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/779,113
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Greta E. No. 5948891and
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33773
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 857 amino acids
; TYPE: amino acid
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; US-08-779-113-2

```

```

alignment_scores:
  Quality: 62.00      Length: 60
  Ratio: 2.000       Gaps: 2
  Percent Similarity: 51.667   Percent Identity: 31.667

```

```

alignment_block:
US-09-049-696-12 x US-08-779-113-2 ..

```

Align seg 1/1 to: US-08-779-113-2 from: 1 to: 857

```

16  CCCAGCCCTCTGTAGTTTATGCAAAATTCGCCAAGGAGCCTCCCAAT 65
   ||| |||:|||||:||||| ||| |||:|||||:|||||
   1 PropheaspLeuValIlePropheAlaValArgLysGly..... 13
66  TCTCAGGCCAGTGTCAAGCCCTGATGAATCAGTGAATGGAACACAG 115
   ||| |||:|||||:||||| ||| |||:|||||:|||||
14  .....GluIleThrGlyGluValHisMetProSerGlyLysThra 27
   ||| |||:|||||:||||| ||| |||:|||||:|||||
116 TTACTTGGAACTACTGATAATGGACGAGGTGCTGATGCTACTAAGGAT 165
   ||| |||:|||||:||||| ||| |||:|||||:|||||
27  lathrProGluIleValAspAsn.....Lysasp 36
   ||| |||:|||||:||||| ||| |||:|||||:|||||
166 GACGGTGTCTACTCAAGGTATTTCACAACT 195
   ||| |||:|||||:||||| ||| |||:|||||:|||||
37  GlyThrValThrValArgTyrAlaProThr 46

```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-583-562B-2

```

seq_documentation_block:
; Sequence 2, Application US/08583562B
; Patent No. 5922570
; GENERAL INFORMATION:
; APPLICANT: Staunton, Donald
; APPLICANT: Harris, Edith
; TITLE OF INVENTION: Cytoplasmic Modulators of Integrin
; TITLE OF INVENTION: Binding
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

TREATMENT

US-09-049-696-12 x US-09-295-028-66

Align seg 1/1 to: US-09-295-028-66 from: 1 to: 456

```
13 TTCCCGACCCCTCGTAGTTTATGCAAAATATCGCCAAGGAGCTCCCC 62
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
258 TyrProGlnHisValAlaValGluAla...ValArgLysGlyAlaAlaAs 273
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
63 AATTCTCAGGCCAGTGTACAGCCCTGATTAATGAAATCAGTGAAT..... 105
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
273 pValLeuArgHisLeuIleThrThrGluValIleSerValAsnGluGluI 290
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
106 .....GGAATA.....ACAGTTACCTTGGAACTACTGGAT 135
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
290 leThrThrProGluGlyLysLysThrThrLeuThrAlaGluAlaLeuThr 306
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
136 AATGGA 141
   :::::
307 SerGly 308
```

seq_name: /cgn2_5/ptodata/2/laa/6B_COMB.pep:US-09-106-582-66

```
seq_documentation_block:
: Sequence 66, Application US/09106582
: Patent No. 6306402
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Lodes, Michael J.
: APPLICANT: Houghton, Raymond
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
: NUMBER OF SEQUENCES: 73
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED and BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/106,582
: FILING DATE: 29-JUN-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Maki, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.439C2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-622-4900
: TELEFAX: 206-682-6031
: INFORMATION FOR SEQ ID NO: 66:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 456 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-106-582-66
```

```
alignment_scores:
  Quality: 61.00      Length: 52
  Ratio: 1.906      Gaps: 3
Percent Similarity: 61.538      Percent Identity: 36.538
```

alignment_block:

US-09-049-696-12 x US-09-106-582-66

Align seg 1/1 to: US-09-106-582-66 from: 1 to: 456

```
13 TTCCCGACCCCTCGTAGTTTATGCAAAATATCGCCAAGGAGCTCCCC 62
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
258 TyrProGlnHisValAlaValGluAla...ValArgLysGlyAlaAlaAs 273
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
63 AATTCTCAGGCCAGTGTACAGCCCTGATTAATGAAATCAGTGAAT..... 105
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
273 pValLeuArgHisLeuIleThrThrGluValIleSerValAsnGluGluI 290
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
106 .....GGAATA.....ACAGTTACCTTGGAACTACTGGAT 135
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
290 leThrThrProGluGlyLysLysThrThrLeuThrAlaGluAlaLeuThr 306
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
136 AATGGA 141
   :::::
307 SerGly 308
```

seq_name: /cgn2_6/ptodata/2/laa/5B_COMB.pep:US-08-752-307B-9

```
seq_documentation_block:
: Sequence 9, Application US/08752307B
: Patent No. 5952171
: GENERAL INFORMATION:
: APPLICANT: McCarthy, Sean A.
: APPLICANT: Gearing, David P.
: APPLICANT: Levinson, Douglas A.
: TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson, P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: US
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/752,307B
: FILING DATE: 19-NOV-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Meiklejohn, Ph.D., Anita L.
: REGISTRATION NUMBER: 35,283
: REFERENCE/DOCKET NUMBER: 09404/020001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-542-5070
: TELEFAX: 617-542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 615 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-752-307B-9
```

```
alignment_scores:
  Quality: 60.50      Length: 65
  Ratio: 1.440      Gaps: 3
Percent Similarity: 64.615      Percent Identity: 29.231
```

alignment_block:

US-09-049-696-12 x US-08-752-307B-9

Align seg 1/1 to: US-08-752-307B-9 from: 1 to: 615

```
31 GTTATGCAATATTCGCAAGAGGCTCCCAATTCCTCAGGCCAGTGT 80
||||| |||:||||| |||:|||||
425 ValtyrLeuAsnValGln.....AlaGluProThrIleSerGluAl 439
81 CACAGCCCTGATTGATCAGTGAATGGAAGAAACAGTTACTTGGAACTAC 130
: ||| :||||:||||:||||:||||:||||:
439 aProAlaAlaValSerThrValAspGlyArgAsnValThrIleLysCysA 456
131 TGGATAATGGAGCAGGTGCTGCTACTAAAGGATGACGGTGCTCTACTCA 180
|||||:|||||:|||||:|||||
456 rgValAsnGlySerProLeuVallys.....TrpLeu 468
181 AGTATTTCACAACTATGACAGCAATGGTAGATACAGTGTAATAA 225
||| :||||:||||:||||:||||:||||:
469 Arg...AlaSerAsnTrpLeuThrGlyGlyArgTyrAsnValGln 482
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-797-366-3

seq_documentation_block:

```
; Sequence 3, Application US/08797366
; Patent No. 5853702
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Christgau, Stephan
; APPLICANT: Halkier, Torben
; APPLICANT: Shuster, Jeff
; APPLICANT: Fuglsang, Claus Crone
; TITLE OF INVENTION: Penicillium Purpurogenum Mutanases
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58537020 No. 5853702disk of No. 5853702th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/797.366
; FILING DATE: 09-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4593.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 630 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-797-366-3
```

alignment_scores:
Quality: 60.00 Length: 64
Ratio: 1.667 Gaps: 4
Percent Similarity: 56.250 Percent Identity: 32.812

alignment_block:

US-09-049-696-12/rev x US-08-797-366-3 ..

Align seg 1/1 to: US-08-797-366-3 from: 1 to: 630

```
219 ACTGTATCTACCATTCCTGTCATAGTTGTGAATACCTTGAGTAGACAC 170
||||| |||:||||| |||:|||||
517 ThrThrSerThrIleSerThrThrPheIleSer...ThrThrTh 532
169 CGTCATCCCTTAGTCAGCATCAGACCTGCTCCATTATCCAGTAGTTCCAAG 120
|||||:|||||:|||||:|||||
532 rThrThrThrSerSerAlaAlaThrSerThrThr.....ThrG 545
119 GFAACTGTTTTTCCATTCCTACTGATTCATCAATCAGGGCTGTGACACATGGCCCT 70
|||||:||||:||||:||||:||||:
545 lYThrCysIleAla.....GlyThrGlyPro 553
69 GAGATTGGGGAGGCTCCTTGGCGAATATTTGCATAAAACTAC 28
|||||:||||:||||:||||:|||||
554 AspAsnTyrSerGly...LeuCysSerPheCysCysAsnTyr 566
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-956-268-3

seq_documentation_block:

```
; Sequence 3, Application US/08956268
; Patent No. 5874275
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Christgau, Stephan
; APPLICANT: Halkier, Torben
; APPLICANT: Shuster, Jeff
; APPLICANT: Fuglsang, Claus Crone
; TITLE OF INVENTION: Penicillium Purpurogenum Mutanases
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58742750 No. 5874275disk of No. 5874275th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956.268
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/797.366
; FILING DATE: 09-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4593.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 630 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-956-268-3
```

alignment_scores:
Quality: 60.00 Length: 64
Ratio: 1.667 Gaps: 4
Percent Similarity: 56.250 Percent Identity: 32.812

alignment_block:

US-09-049-696-12/rev x US-08-956-268-3 ..

Align seg 1/1 to: US-08-956-268-3 from: 1 to: 630

```
219 ACTGATCTACACATCGGTGTCATAGTTGTGAATAACCTTGAGTAGACAC 170
|||||
517 ThrThrSerThrThrThrThrThrThrThrThrThrThrThrThrThrThr 532
|||||
169 CGTCATCTCTAGTACATCAGCAGCTGCTCCATTCATTCAGTAGTCCCAAG 120
|||||
532 ThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 545
|||||
119 GTAACGTGTTTTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 70
|||||
545 LysThrCysIleAla.....GlyThrGlyPro 553

59 GAGAAATCGGGAGGCTCTTCGCGAATATTCGATTAACACTAC 28
|||||
554 AspAsnTyrSerGly...LeuCysSerPheCysCysAsnTyr 566
|||||
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-094-557-3

```
seq_documentation_block:
; Sequence 3, Application US/09094557
; Patent No. 6225531
; GENERAL INFORMATION:
; APPLICANT: Kakitani, Makoto
; APPLICANT: Umamoto, Naoyuki
; APPLICANT: Ishida, Isao
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Yoshikawa, Masaaki
; APPLICANT: Yamoka, Naoto
; TITLE OF INVENTION: GLUCAN ELICITOR RECEPTOR, DNA MOLECULE
; CODING THEREFOR, FUNGUS-RESISTANT PLANTS TRANSFORMED WITH
; THE DNA MOLECULE AND METHOD FOR CREATING THE PLANTS
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/094,557
; FILING DATE: 15-JUN-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03653
; FILING DATE: 13-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 136100/1994
; FILING DATE: 17-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 347823/1995
; FILING DATE: 15-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/591,566
; FILING DATE: 14-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 081356/0116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 347 amino acids
; TYPE: amino acid
```

```
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-094-557-3

alignment_scores:
Quality: 59.50 Length: 81
Ratio: 1.266 Gaps: 3
Percent Similarity: 58.025 Percent Identity: 29.630

alignment_block:
US-09-049-696-12 x US-09-094-557-3

Align seg 1/1 to: US-09-094-557-3 from: 1 to: 347

1 GACACCCAGCAAAATCCCCAGCCCTCTGCTAGTTATGCAAAATATTCGCCA 50
|||||
223 AspTyrAlaLeuPheArgSerProSerValValValGlnAspGlySerLe 239
|||||
51 AGGAGCCTCCCAATTCAGGGCCAGTGTACAGCCCTG.....A 91
|||||
239 uGlyTyrArgAsnLeuPheAspAlaSerValAspAlaValTyrAlaAlaL 256
|||||
92 TTGAATCAGTGAATGGAACACAGTTACCTTGGAACTACTGGATAATGGA 141
|||||
256 euGluLysAlaGlyGlySerLeuAsnIleValValSerGluSerGly 272
|||||
142 .....GCAGGTGCTGATGCTACTAAGGATGACGGTGTCTACTCAAG 182
|||||
273 TrpProSerSerGlyGlyThrAlaThrSerLeuAspAsnAla..... 286
|||||
183 GTATTTCACACTTATGACACGAATGGTAGATACAGTGTATAAA 225
|||||
287 .....ArgThrTyrAsnThrAsnLeuValArgAsnValLys 298
|||||

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US95-00362-5

seq_documentation_block:
; Sequence 5, Application PC/TUS9500362
; GENERAL INFORMATION:
; APPLICANT: IMMUNEX CORPORATION
; TITLE OF INVENTION: Ligand That Binds Fas Antigen
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: US
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/00362
; FILING DATE: 06-JAN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/179,138
; FILING DATE: 07-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/190,559
; FILING DATE: 01-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Anderson, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2805-WO
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 amino acids
; TYPE: amino acid
```


OM of: US-09-049-696-5 to: SPTRMBL_17:* out_format : pfs

Date: Mar 30, 2002 2:46 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frame+ntp.model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09049696/runat_28032002_145238_2053/app_query.fasta_1.12579
-DB=SPTRMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-CAPEXT=4.000 -MINMATCH=0.100 -LOOCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-XGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1_705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-5
Query length: 220
Database: SPTRMBL_17:*
Database sequences: 473505
Database length: 146272329
Search time (sec): 805.760000

score_list:

Sequence	Strd	Orig	Zscore	EScore	Len	Documentation
sp_human:095151	+	386.00	858.11	1.7e-40	914	! 095151 homo sapiens (human).
sp_human:09UPC6	+	386.00	858.11	1.7e-40	914	! 09UPC6 homo sapiens (human).
sp_human:09UNF6	+	386.00	858.11	1.7e-40	914	! 09UNF6 homo sapiens (human).
sp_mammal:09TRUB5	+	357.00	791.99	8.4e-37	917	! 09TRUB5 sus scrofa (pig).
sp_rodent:098826	+	343.00	760.11	5.0e-35	913	! 098826 mus musculus (mouse).
sp_rodent:09B726	+	343.00	760.11	5.0e-35	913	! 09B726 mus musculus (mouse).
sp_human:09UNF7	+	284.00	625.59	1.6e-27	917	! 09UNF7 homo sapiens (human).
sp_rodent:098860	+	259.00	568.73	2.3e-24	901	! 098860 mus musculus (mouse).
sp_rodent:09X015	+	258.00	568.72	2.3e-24	902	! 09X015 mus musculus (mouse).
sp_rodent:09R070	+	258.00	566.44	3.1e-24	902	! 09R070 mus musculus (mouse).
sp_rodent:09B0R4	+	249.00	552.56	4.9e-23	902	! 09B0R4 mus musculus (mouse).
sp_mammal:018744	+	249.00	546.58	4.4e-23	905	! 018744 bos taurus (bovine).
sp_mammal:018743	+	249.00	545.90	4.4e-23	905	! 018743 bos taurus (bovine).
sp_mammal:018741	+	249.00	545.90	4.4e-23	905	! 018741 bos taurus (bovine).
sp_human:09UOC9	+	232.00	506.87	6.2e-21	943	! 09UOC9 homo sapiens (human).
sp_human:09Y6N2	+	232.00	506.87	6.2e-21	943	! 09Y6N2 homo sapiens (human).
sp_human:09Y6N3	+	75.00	157.76	0.6234	262	! 09Y6N3 homo sapiens (human).
sp_rodent:063394	-	75.00	156.54	0.6099	313	! 063394 rattus norvegicus (rat).
sp_plant:09LDB3	-	69.50	140.08	2.84	556	! 09LDB3 arabidopsis thaliana (mc
sp_plant:09LQV4	+	69.00	140.71	3.39	429	! 09LQV4 arabidopsis thaliana (mc
sp_human:095226	-	69.00	130.49	2.82	1912	! 095226 homo sapiens (human).
sp_human:09UHB1	-	69.00	130.26	2.81	1977	! 09UHB1 homo sapiens (human).
sp_rodent:09JIS7	+	69.00	130.24	2.81	1985	! 09JIS7 mus musculus (mouse).
sp_human:09BUQ7	+	67.00	135.63	6.03	463	! 09BUQ7 homo sapiens (human).
sp_bacteria:09A4U5	+	66.00	138.57	8.87	216	! 09A4U5 caulobacter crescentus.
sp_human:09UH39	-	66.00	135.40	8.38	343	! 09UH39 homo sapiens (human).
sp_bacteria:0998D7	+	65.00	131.36	10.87	444	! 0998D7 staphylococcus aureus su
sp_vertebrate:09PU26	+	64.50	131.05	12.77	393	! 09PU26 trachemys scripta elegans
sp_virus:065482	-	64.00	140.15	17.77	88	! 065482 berne virus (bev).
sp_mammal:029282	+	64.00	138.68	17.31	109	! 029282 sus scrofa (pig).
sp_mammal:09NSN3	+	64.00	135.03	16.21	186	! 09NSN3 homo sapiens (human).
sp_fungi:09P7V4	+	64.00	132.21	15.40	281	! 09P7V4 schizosaccharomyces pombe
sp_rodent:09ER66	+	64.00	130.63	14.97	354	! 09ER66 mus musculus (mouse).
sp_rodent:09B9N3	-	64.00	129.95	14.79	391	! 09B9N3 mus musculus (mouse).
sp_human:09NZN0	+	64.00	126.39	13.87	658	! 09NZN0 homo sapiens (human).
sp_human:09NP60	+	64.00	126.11	13.80	686	! 09NP60 homo sapiens (human).
sp_rodent:09ERS6	+	64.00	126.11	13.80	686	! 09ERS6 mus musculus (mouse).
sp_bacteria:006459	+	64.00	122.31	12.89	1195	! 006459 thermus aquaticus (subs
sp_invertebrate:09UIU2	-	63.50	129.61	17.37	348	! 09UIU2 caenorhabditis elegans

seq_name: sp_human:095151

seq_documentation_block:

ID 095151 PRELIMINARY; PRT; 914 AA.

AC 095151;

DT 01-MAY-1999 (TREMBLREL, 10, Created)

DT 01-MAY-1999 (TREMBLREL, 10, Last sequence update)

DE 01-JUN-2001 (TREMBLREL, 17, Last annotation update)

DE CALCITRIUM-DEPENDENT CHLORIDE CHANNEL-1.

GN HCLCAL.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCAL_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=SMALL INTESTINE;

RX MEDLINE=99047526; PubMed=9828122;

RA Gruber A.D., Eble R.C., Ji H.L., Schreur K.D., Fuller C.M.,

KA Paull B.O.;

RT "Genomic cloning, molecular characterization, and functional analysis

RT of human CLCAL, the first human member of the family of Ca2+-activated

RL Genomics 54:200-214(1998).

DR EMBL: AF039400; AAC95428.1;

DR InterPro: IPR0010131; ATPase.gamma.

DR InterPro: IPR002035; VWFA.

DR PROSITE: PS00153; ATPASE_GAMMA; UNKNOWN_1.

DR PROSITE: PS00234; VWFA; 1.

DR SMART: SM00327; VWA; 1.

SQ SEQUENCE 914 AA; 100174 MW; 7066440C46557FA3 CRC64;

alignment_scores:

Quality: 386.00 Length: 72

Ratio: 5.361 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-5 x 095151

Align seg 1/1 to: 095151 from: 1 to: 914

3 ATAGTTGAATCTGTACAGACAAACACCAACCAAGAGCTCCAAACAA 52

246 lleValGluPheCysThrGluGlnAsnHisAsnLysGluAlaProAsnly 262

53 CCAAAATCAAAATCAATCTCCGAGGACATCGGGAAGTGATCCGTGATT 102

262 SGLNAsnGlnLysCysAsnLeuArgSerThrTrpGluValIleArgAspS 279

103 CTGAGGACTTTTAAAGAAAACCACTCTATGACACACAGCCACCAATGCC 152

279 erGluAspPheLysLysThrThrProMetThrThrGlnProProAsnPro 295

153 ACCCTCTCATTCCTCCATTCGACATTCGACAAAGAAATTTGTGTTTACTCTTGA 202

296 ThrPheSerLeuLeuGlnIleGlyGlnArgIleValCysLeuValLeuAs 312

203 CAAATCTCGAAGCATG 218

312 pLysSerGlySerMet 317

seq_name: sp_human:09UPC6

seq_documentation_block:

ID 09UPC6 PRELIMINARY; PRT; 914 AA.

```
AC Q9UPC6; 2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL1
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95047526; PubMed=9828122;
RA Gruber A.D., Elbie R.C., Ji H.L., Schreuer K.D., Fuller C.M.,
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
RT of human CLCAL1, the first human member of the family of Ca2+-activated
RT Cl-channel proteins."
RL Genomics 54:200-214(1998).
DR EMBL; AF039401; AAC95429.1; -
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 914 AA; 100160 MW; 7761362B66752AA3 CRC64;

alignment_scores:
  Quality: 386.00      Length: 72
  Ratio: 5.361        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-5 x Q9UPC6
Align seg 1/1 to: Q9UPC6 from: 1 to: 914

3 ATAGTTGAATTCGTGACAGAACAAACCAACAAAGAGCTCCAAACAA 52
|||||
246 lIeValGluPheCysThrGluGlnAsnHisAsnLysGluAlaProAsnLy 262

53 GCAAAATCAAAATGCAATCTCCGAGCACATGGGAAGTGCCTGATT 102
|||||
262 sGlnAsnGlnLysCysAsnLeuArgSerThrTrpGluValIleArgAspS 279

103 CTGAGGACTTTAAGAAACCACTCTATGACACACAGCCACCAATCCC 152
|||||
279 erGluAspPheLysThrThrPrometThrThrGlnProProAsnPro 295

153 ACCTTCATGTCGAGATGGACAAGAATTGTGTGTTAGTCTTGA 202
|||||
296 ThrPheSerLeuLeuGlnIleGlyGlnArgIleValCysLeuValLeuAs 312

203 CAAATCTGGAGCATG 218
|||||
312 pLysSerGlySerMet 317

seq_name: sp_human:Q9UPC6

seq_documentation_block:
ID Q9UPC6 PRELIMINARY; PRT; 914 AA.
AC Q9UPC6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 1.
GN CACCL1
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX TISSUE=SMALL INTESTINE, AND COLON;
RX MEDLINE=99364503; PubMed=10437792;
RA Aguel M., Vermaat T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
RT chloride channel (CaCC) family predominantly expressed in the
RT digestive tract and trachea."
RL PDBS Lett. 455:295-301(1999).
DR EMBL; AF127036; AAD25487.1; -
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 914 AA; 100202 MW; B58B451D97887D6B CRC64;

alignment_scores:
  Quality: 386.00      Length: 72
  Ratio: 5.361        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-5 x Q9UNF6
Align seg 1/1 to: Q9UNF6 from: 1 to: 914

3 ATAGTTGAATTCGTGACAGAACAAACCAACAAAGAGCTCCAAACAA 52
|||||
246 lIeValGluPheCysThrGluGlnAsnHisAsnLysGluAlaProAsnLy 262

53 GCAAAATCAAAATGCAATCTCCGAGCACATGGGAAGTGCCTGATT 102
|||||
262 sGlnAsnGlnLysCysAsnLeuArgSerThrTrpGluValIleArgAspS 279

103 CTGAGGACTTTAAGAAACCACTCTATGACACACAGCCACCAATCCC 152
|||||
279 erGluAspPheLysThrThrPrometThrThrGlnProProAsnPro 295

153 ACCTTCATGTCGAGATGGACAAGAATTGTGTGTTAGTCTTGA 202
|||||
296 ThrPheSerLeuLeuGlnIleGlyGlnArgIleValCysLeuValLeuAs 312

203 CAAATCTGGAGCATG 218
|||||
312 pLysSerGlySerMet 317

seq_name: sp_mammal:Q9TUB5

seq_documentation_block:
ID Q9TUB5 PRELIMINARY; PRT; 917 AA.
AC Q9TUB5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN.
GN AECC.
OS Sus scrofa (Pig)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=ILEAL MUCOSA;
RX Gaspar K.J., Gabriel S.E., Racette K.J., Forsyth G.W.;
RT "A cdna involved in porcine exocrine chloride conductance."
RT Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF095584; AAF00077.1; -
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
```


SQ SEQUENCE 917 AA; 100735 MW; 649F41944F7EAD19 CRC64;

alignment_scores:
Quality: 357.00 Length: 72
Ratio: 5.174 Gaps: 0
Percent Similarity: 95.833 Percent Identity: 91.667

alignment_block:

US-09-049-696-5 x Q9TUB5

Align seg 1/1 to: Q9TUB5 from: 1 to: 917

3 ATAGTTCAATTCGTACAGAACAAACCAACAAAGAGCTCCAAACAA 52
:::|||||
246 ValValGluPheCysThrGluLysAsnHisAsnGlnGluAlaProAsnAs 262
53 GCAAAATCAAAATGCAATCTCCGAAGCACATCGGAAGTGCCTGATT 102
|||||
262 pGlnAsnGlnLysCysAsnLeuArgSerThrTrpGluValIleGlnAspS 279
103 CTGAGGACTTTAAGAAACCACTCTATGACACACAGCACCACCAATCCC 152
|||||
279 erGluAspPheLysThrThrPrometThrThrGlnProAlaPro 295
153 ACCTTCTCATTCGTCAGATTGGACAAGAATGTGCTGTTAGTCCTTGA 202
|||||
296 ThrPheSerLeuLeuGlnIleGlyGlnArgIleValCysLeuValLeuAs 312
203 CAAATCTGGAAGCATG 218
|||||
312 pLysSerGlySerMet 317

seq_name: sp_rodent:O88826

seq_documentation_block:

ID O88826 PRELIMINARY; PRT; 913 AA.
AC O88826;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GOB-5 PROTEIN.
GN CLCA3 OR GOB-5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RA Komiyama T., Tanigawa Y., Hirohashi S.;
RT "Cloning of a gene, gob-5, which is expressed in intestinal goblet
cells in mice."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017156; BAA33743.1; -.
DR MGD; MGI:1346342; Clca3.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 913 AA; 100070 MW; A7FA2F9E1089806D CRC64;

alignment_scores:

Quality: 343.00 Length: 72
Ratio: 4.900 Gaps: 0
Percent Similarity: 97.222 Percent Identity: 86.111

alignment_block:

US-09-049-696-5 x O88826

Align seg 1/1 to: O88826 from: 1 to: 913

3 ATAGTTGAATTCGTACAGAACAAACCAACAAAGAGCTCCAAACAA 52
:::|||||
247 ValValGluPheCysThrGluLysAsnHisAsnGlnGluAlaProAsnAs 263
53 GCAAAATCAAAATGCAATCTCCGAAGCACATCGGAAGTGCCTGATT 102
|||||
263 pGlnAsnGlnLysCysAsnLeuArgSerThrTrpGluValIleGlnGluS 280
103 CTGAGGACTTTAAGAAACCACTCTATGACACACAGCACCACCAATCCC 152
|||||
280 erGluAspPheLysThrThrPrometThrAlaGlnProAlaPro 296
153 ACCTTCTCATTCGTCAGATTGGACAAGAATGTGCTGTTAGTCCTTGA 202
|||||
297 ThrPheSerLeuLeuGlnIleGlyGlnArgIleValCysLeuValLeuAs 313
203 CAAATCTGGAAGCATG 218
|||||
313 pLysSerGlySerMet 318

seq_name: sp_rodent:Q9D726

seq_documentation_block:

ID Q9D726 PRELIMINARY; PRT; 913 AA.
AC Q9D726;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CHLORIDE CHANNEL CALCIUM ACTIVATED 3.
GN CLCA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=STOMACH;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Seito T., Okazaki Y., Gojobori T., Hono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Yashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK008659; BAB25815.1; -.
DR MGD; MGI:1346342; Clca3.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 913 AA; 100111 MW; 491E8584B06D9A89 CRC64;

alignment_scores:

Quality: 343.00 Length: 72
Ratio: 4.900 Gaps: 0
Percent Similarity: 97.222 Percent Identity: 86.111

[illegible][illegible]

```

seq_documentation_block:
ID Q9QX15 PRELIMINARY; PRT; 902 AA.
AC Q9QX15;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIUM-SENSITIVE CHLORIDE CONDUCTANCE PROTEIN-1.
GN CLCA1 OR MCLCA1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1];
RP SEQUENCE FROM N.A.
RC TISSUE=WHOLE LUNG;
RX MEDLINE=99041980; PubMed=9822685;
RA Gandhi R., Elble R.C., Gruber A.D., Schreuer K.D., Ji H.-L.,
RA Fuller C.M., Pauli B.U.;
RT "Molecular and functional characterization of a calcium-sensitive
RT chloride channel from mouse lung.";
RL J. Biol. Chem. 273:32096-32101(1998).
DR EMBL; AF047838; AAC79982.1; -
DR MGD; MGI:1316732; Clcal.
DR InterPro; IPR002035; vWFA.
DR SMART; SM00327; vWA; 1.
DR PROSITE; PS50234; vWFA; 1.
SQ SEQUENCE 902 AA; 100107 MW; D9A915E698EA2D85 CRC64;

alignment_scores:
Quality: 259.00 Length: 74
Ratio: 4.317 Gaps: 1
Percent Similarity: 81.081 Percent Identity: 66.216

alignment_block:
US-09-049-696-5 x Q9QX15
Align seg 1/1 to: Q9QX15 from: 1 to: 902

3 ATAGTTGAATTCGTACAGAACAAACACAAAGAGCTCCAAACAA 52
:::|||||
246 ValValGluPheCysThrGluLysAsnHisAsnAlaGluAlaProAsnLe 262

53 GCAAAATCAAAATGCAATCTCCGAGACATGGAAGTGTGTTAGT 102
|||||::: ||||| |||||||
262 uGlnAsnLysMetCysAsnArgSerThrTrpAspValIleLysThrS 279

103 CTGAGGACTTTAAGAAACCACTCTATG.....ACAACACGCCACCA 146
|| |||||
279 erAlaAspPheGlnAsnSerProMetArgGlyThrGluAlaProPro 295

147 AATCCACCTTCTCATTCGTCGAGATTGGACAAAGAAATGTGTTAGT 196
|||||
296 ProProThrPheSerLeuLeuLysSerArgArgValValCysLeuVa 312

197 CTTGACAAATCTGGAAGCATG 218
|||||
312 lLeuAspLysSerGlySerMet 319

seq_name: sp_rodent:Q9R070

seq_documentation_block:
ID Q9R070 PRELIMINARY; PRT; 902 AA.
AC Q9R070;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CA(2+)-SENSITIVE CHLORIDE CHANNEL 2.
GN CLCA2 OR CAC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

OX NCBI_TaxID=10090;
RN [1];
RP SEQUENCE FROM N.A.
RC TISSUE=MAMARY;
RX MEDLINE=20012773; PubMed=10544033;
RA Lee D., Ha S., Kho Y., Kim J., Cho K., Baik M., Choi Y.;
RT "Induction of mouse Ca(2+)-sensitive chloride channel 2 gene during
RT involution of mammary gland.";
RL Biochem. Biophys. Res. Commun. 264:933-937(1999).
DR EMBL; AF108501; AAF12731.1; -
DR MGD; MGI:1931471; Clca2.
DR InterPro; IPR002035; vWFA.
DR SMART; SM00327; vWA; 1.
DR PROSITE; PS50234; vWFA; 1.
SQ SEQUENCE 902 AA; 99866 MW; 0FB746CF5FB3D07F CRC64;

alignment_scores:
Quality: 258.00 Length: 74
Ratio: 4.300 Gaps: 1
Percent Similarity: 81.081 Percent Identity: 66.216

alignment_block:
US-09-049-696-5 x Q9R070
Align seg 1/1 to: Q9R070 from: 1 to: 902

3 ATAGTTGAATTCGTACAGAACAAACACAAAGAGCTCCAAACAA 52
:::|||||
246 ValValGluPheCysThrGluAsnAsnHisAsnAlaGluAlaProAsnLe 262

53 GCAAAATCAAAATGCAATCTCCGAGACATGGAAGTGTGTTAGT 102
|||||::: ||||| |||||||
262 uGlnAsnLysMetCysAsnArgSerThrTrpAspValIleLysAlaS 279

103 CTGAGGACTTTAAGAAACCACTCTATG.....ACAACACGCCACCA 146
|| |||||
279 erAlaAspPheGlnAsnSerProMetArgGlyThrGluAlaProPro 295

147 AATCCACCTTCTCATTCGTCGAGATTGGACAAAGAAATGTGTTAGT 196
|||||
296 ProProThrPheSerLeuLeuLysSerArgArgValValCysLeuVa 312

197 CTTGACAAATCTGGAAGCATG 218
|||||
312 lLeuAspLysSerGlySerMet 319

seq_name: sp_rodent:Q9EQR4

seq_documentation_block:
ID Q9EQR4 PRELIMINARY; PRT; 902 AA.
AC Q9EQR4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ENDOTHELIAL CHLORIDE CHANNEL.
GN CLCA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=C57/BL6;
RA Elble R.C., Pauli B.;
RT "A murine endothelial homolog of Lu-ECAM-1 associated with lung
RT melanoma metastasis.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF115852; AAG47626.1; -
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR002035; vWFA.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00327; vWA; 1.

```

DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 902 AA; 99855 MW; F2BA57DAA5D4AAAD CRC64;

alignment_scores:
Quality: 258.00 Length: 74
Ratio: 4.300 Gaps: 1
Percent Similarity: 81.081 Percent Identity: 66.216

alignment_block:
US-09-049-696-5 x Q9EQR4

Align seg 1/1 to: Q9EQR4 from: 1 to: 902

```
3 ATAGTTGAATTCGTACAGAAACCAACCAACCAAGAGCTCCAAACAA 52
: : : : : : : : : : : : : : : : : : : : : : : : : :
246 ValValGluPheCysThrGluAsnHisAsnAlaGluAlaProAsnLe 262
53 GCAAAATCAAAATGCAATCTCCGAGACATGGGAAGTGCCTGATT 102
: : : : : : : : : : : : : : : : : : : : : : : : : :
262 uGlnAsnLysMetCysAsnArgSerThrTrpAspValIleTysAla 279
103 CTGAGGACTTTTAAAGAAACCACTCTCTATG.....ACAACACAGCCACCA 146
: : : : : : : : : : : : : : : : : : : : : : : : : :
279 erValAspPheGlnAsnSerProMetArgGlyThrGluAlaProPro 295
147 AATCCACCTTCTCATTCGTGCAGATTGGACAAAGAATTGTGTGTAGT 196
: : : : : : : : : : : : : : : : : : : : : : : : : :
296 ProProThrPheSerLeuLeuLysSerArgArgValValCysLeuVa 312
197 CTTGACAAATCTGGAGCATG 218
: : : : : : : : : : : : : : : : : : : : : : : : : :
312 lLeuAspLysSerGlySerMet 319
```

seq_name: sp_mammal:O18744

seq_documentation_block:
ID O18744 PRELIMINARY; PRT; 342 AA.

```
AC 018742;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001264; AAB86532.1; -
SQ SEQUENCE 342 AA; 39037 MW; 6A491480B4BED01A CRC64;
```

alignment_scores:
Quality: 249.00 Length: 74
Ratio: 4.016 Gaps: 1
Percent Similarity: 83.784 Percent Identity: 63.514

alignment_block:
US-09-049-696-5 x O18744

Align seg 1/1 to: O18744 from: 1 to: 342

```
3 ATAGTTGAATTCGTACAGAAACCAACCAACCAAGAGCTCCAAACAA 52
: : : : : : : : : : : : : : : : : : : : : : : : : :
247 ValThrGluPheCysThrGluLysThrHisAsnThrGluAlaProAsnLe 263
53 GCAAAATCAAAATGCAATCTCCGAGACATGGGAAGTGCCTGATT 102
```

```
263 uGlnAsnLysMetCysAsnGlySerThrTrpAspValIleMetAsnS 280
: : : : : : : : : : : : : : : : : : : : : : : : : :
103 CTGAGGACTTTTAAAGAAACCACTCTCTATGACA.....ACACAGCCACCA 146
: : : : : : : : : : : : : : : : : : : : : : : : : :
280 erValAspPheGlnAsnThrSerProMetThrGluMetAsnProProThr 296
147 AATCCACCTTCTCATTCGTGCAGATTGGACAAAGAATTGTGTGTAGT 196
: : : : : : : : : : : : : : : : : : : : : : : : : :
297 HisProThrPheSerLeuLeuLysSerLysGlnArgValValCysLeuVa 313
197 CTTGACAAATCTGGAGCATG 218
: : : : : : : : : : : : : : : : : : : : : : : : : :
313 lLeuAspLysSerGlySerMet 320
```

seq_name: sp_mammal:O18742

seq_documentation_block:
ID O18742 PRELIMINARY; PRT; 794 AA.

```
AC 018742;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001262; AAB86530.1; -
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWFA; 1.
SQ SEQUENCE 794 AA; 88509 MW; B695E7256FC2C632 CRC64;
```

alignment_scores:
Quality: 249.00 Length: 74
Ratio: 4.016 Gaps: 1
Percent Similarity: 83.784 Percent Identity: 63.514

alignment_block:
US-09-049-696-5 x O18742

Align seg 1/1 to: O18742 from: 1 to: 794

```
3 ATAGTTGAATTCGTACAGAAACCAACCAACCAAGAGCTCCAAACAA 52
: : : : : : : : : : : : : : : : : : : : : : : : : :
247 ValThrGluPheCysThrGluLysThrHisAsnThrGluAlaProAsnLe 263
53 GCAAAATCAAAATGCAATCTCCGAGACATGGGAAGTGCCTGATT 102
: : : : : : : : : : : : : : : : : : : : : : : : : :
263 uGlnAsnLysMetCysAsnGlySerThrTrpAspValIleMetAsnS 280
103 CTGAGGACTTTTAAAGAAACCACTCTCTATGACA.....ACACAGCCACCA 146
: : : : : : : : : : : : : : : : : : : : : : : : : :
280 erValAspPheGlnAsnThrSerProMetThrGluMetAsnProProThr 296
147 AATCCACCTTCTCATTCGTGCAGATTGGACAAAGAATTGTGTGTAGT 196
: : : : : : : : : : : : : : : : : : : : : : : : : :
297 HisProThrPheSerLeuLeuLysSerLysGlnArgValValCysLeuVa 313
197 CTTGACAAATCTGGAGCATG 218
: : : : : : : : : : : : : : : : : : : : : : : : : :
313 lLeuAspLysSerGlySerMet 320
```

seq_name: sp_mammal:O18743

seq_documentation_block:
ID O18743 PRELIMINARY; PRT; 820 AA.
AC O18743;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001263; AAB86531.1; -;
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 820 AA; 91464 MW; 4E01330AEF5A1494 CRC64;

alignment_scores:
Quality: 249.00 Length: 74
Ratio: 4.016 Gaps: 1
Percent Similarity: 83.784 Percent Identity: 63.514
alignment_block:
US-09-049-696-5 x O18743 ..
Align seg 1/1 to: O18743 from: 1 to: 820
3 ATAGTTGAATTCGTACAGAACAAACACAAAGAGCTCCAAACAA 52
247 ValThrGluPheCysThrGluLysThrHisAsnThrGluAlaProAsnLe 263
53 GCAAAATCAAAATGCAATCTCCGAGCAGCATGGGAGTGTGTTAGT 102
263 uGlnAsnLysMetCysAsnGlyLysSerThrTrpAspValIleMetAsnS 280
103 CTGAGGACTTTAAGAAACACCTCTATGACA.....ACACAGCCACCA 146
280 erValAspPheGlnAsnThrSerProMetThrGluMetAsnProProThr 296
147 AATCCACCTTCTCATTGCTGCAGATTGGACAAAGAAATGTGTTAGT 196
297 HisProThrPheSerLeuLeuLysSerLysGlnArgValValCysLeuVa 313
197 CCTTGACAAATCTGGAAGCATG 218
313 lLeuAspLysSerGlySerMet 320

seq_name: sp_mammal:O18741
seq_documentation_block:
ID O18741 PRELIMINARY; PRT; 905 AA.
AC O18741;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;

RA Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001261; AAB86529.1; -;
DR InterPro; IPR002035; VWFA.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF00041; fn3; 1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 905 AA; 101005 MW; D4CA6E7919115E88 CRC64;
alignment_scores:
Quality: 249.00 Length: 74
Ratio: 4.016 Gaps: 1
Percent Similarity: 83.784 Percent Identity: 63.514
alignment_block:
US-09-049-696-5 x O18741 ..
Align seg 1/1 to: O18741 from: 1 to: 905
3 ATAGTTGAATTCGTACAGAACAAACACAAAGAGCTCCAAACAA 52
247 ValThrGluPheCysThrGluLysThrHisAsnThrGluAlaProAsnLe 263
53 GCAAAATCAAAATGCAATCTCCGAGCAGCATGGGAGTGTGTTAGT 102
263 uGlnAsnLysMetCysAsnGlyLysSerThrTrpAspValIleMetAsnS 280
103 CTGAGGACTTTAAGAAACACCTCTATGACA.....ACACAGCCACCA 146
280 erValAspPheGlnAsnThrSerProMetThrGluMetAsnProProThr 296
147 AATCCACCTTCTCATTGCTGCAGATTGGACAAAGAAATGTGTTAGT 196
297 HisProThrPheSerLeuLeuLysSerLysGlnArgValValCysLeuVa 313
197 CCTTGACAAATCTGGAAGCATG 218
313 lLeuAspLysSerGlySerMet 320

This Page Blank (uspto)

OM of: US-09-049-696-5 to: SwissProt_39.* out_format : pfs

Date: Mar 30, 2002 2:51 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+nt2p.model -DEV=xlp
-O=/cgn2_1/USPRO_spool/US09049696/runat_28032002_145238_2085/app_query.fasta_1.12579
-DB=SwissProt_39 -QFMT=fastan -SUFFIX=rs -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGNL_1_165 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-5

Query length: 220

Database: SwissProt_39.*

Database sequences: 100059

Database length: 3664827

Search time (sec): 306.030000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
SwissProt_39:ECIC_BOVIN	+	248.00	524.34	1.7e-22	903
SwissProt_39:IRAE_MOUSE	+	71.50	126.12	0.7691	3079
SwissProt_39:APSA_EMEI	+	69.50	126.46	1.35	1676
SwissProt_39:CCAF_HUMAN	+	69.00	124.12	1.56	1966
SwissProt_39:DPOL_HUMAN	+	66.50	124.92	3.16	875
SwissProt_39:CCAS_HUMAN	+	64.00	113.48	6.40	1873
SwissProt_39:RBL2_MOUSE	+	63.50	116.29	7.37	1135
SwissProt_39:O2B2_HUMAN	+	63.00	124.20	8.49	357
SwissProt_39:KAD_PARDE	+	62.50	126.98	9.77	217
SwissProt_39:COMM_DROME	+	62.00	121.72	11.26	370
SwissProt_39:SUUV3_YEAST	+	62.00	116.35	11.25	737
SwissProt_39:CCAS_HSVB	+	62.00	109.08	11.25	1873
SwissProt_39:VG48_HSVB	+	61.50	114.64	12.96	797
SwissProt_39:O2B3_HUMAN	+	61.00	120.82	14.93	313
SwissProt_39:Y147_HAEM	+	61.00	115.33	14.93	633
SwissProt_39:NFC2_HUMAN	+	61.00	112.38	14.93	925
SwissProt_39:CCAD_MESAU	+	61.00	108.06	14.92	1610
SwissProt_39:CCAD_RAT	+	61.00	105.61	14.92	2203
SwissProt_39:RBL2_HUMAN	+	60.50	109.65	17.19	1139
SwissProt_39:VWF_HUMAN	+	60.50	102.61	17.18	2813
SwissProt_39:POLG_HCAV	+	60.50	100.06	17.18	3898
SwissProt_39:YK69_CAEEL	+	60.00	122.89	19.80	181
SwissProt_39:VGLL_HSV2H	+	60.00	121.23	19.80	224
SwissProt_39:CCG3_HUMAN	+	60.00	118.57	19.80	315
SwissProt_39:CCAC_RAT	+	60.00	103.52	19.79	2171
SwissProt_39:CCAC_HCVB	+	60.00	103.35	19.79	2221
SwissProt_39:POLG_HCVB	+	60.00	98.96	19.79	3898
SwissProt_39:RS27_METUA	+	59.50	130.14	22.81	62
SwissProt_39:CCSA_SPIOL	+	59.50	117.28	22.80	323
SwissProt_39:POC2_HUMAN	+	59.50	116.78	22.80	344
SwissProt_39:IRK9_HUMAN	+	59.50	115.75	22.80	393
SwissProt_39:IRK9_MOUSE	+	59.50	115.75	22.80	393
SwissProt_39:IRK9_RAT	+	59.50	115.75	22.80	393
SwissProt_39:MDM2_MESAU	+	59.50	114.42	22.80	466
SwissProt_39:MSPI_PLAF3	+	59.50	104.41	22.80	1682
SwissProt_39:MSPI_PLAFM	+	59.50	104.33	22.80	1701
SwissProt_39:MSPI_PLAFM	+	59.50	104.33	22.80	1701
SwissProt_39:MSPI_PLAFM	+	59.50	104.21	22.80	1726
SwissProt_39:Y103_SYNV3	+	59.00	114.13	26.26	420

SwissProt_39:RBB8_HUMAN + 59.00 108.21 26.26 897 ! Q99708 homo sapiens (huma
SwissProt_39:ITAE_MOUSE - 59.00 106.16 26.26 1167 ! Q50677 mus musculus (mou
SwissProt_39:CCAC_MOUSE - 59.00 101.44 26.25 2139 ! Q01815 mus musculus (mou
SwissProt_39:CCAC_RAT - 59.00 101.33 26.25 2169 ! P22002 rattus norvegicus (ch
SwissProt_39:CCAD_CHICK - 59.00 101.25 26.25 2190 ! O73700 gallus gallus (ch

seq_name: SwissProt_39:ECIC_BOVIN

seq_documentation_block:

ID ECIC_BOVIN STANDARD; PRT; 903 AA.
AC P54281;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN (CALCIUM-ACTIVATED CHLORIDE
DE CHANNEL).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
PC TISSUE=Trachea;
RX MEDLINE=96125078; PubMed=8537359;
RA Cunningham S.A., Awayda M.S., Bublief J.K., Ismailov I.I.,
RA Arrate M.P., Berdiev B.K., Benos D.J., Fuller C.M.;
RT "Cloning of an epithelial chloride channel from bovine trachea";
RL J. Biol. Chem. 270:31016-31026(1995).
CC !- FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE
CC SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME;
CC MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND
CC TRANSEPITHELIAL TRANSPORT.
CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC !- TISSUE SPECIFICITY: TRACHEA.
CC !- PTM: PHOSPHORYLATED BY CAM-KINASE II.
CC !- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U36445; AAC48511.1; -
CC InterPro: IPR002035; VWFA.
CC SMART: SM00327; VWFA; 1.
CC PROSITE: PS50234; VWFA; 1.
CC Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW Calcium channel; Phosphorylation;
FT TRANSMEM 7 27
FT TRANSMEM 331 351
FT TRANSMEM 617 637
FT TRANSMEM 883 903
FT DOMAIN 308 476
FT CARBOHYD 75 75
FT CARBOHYD 278 278
FT CARBOHYD 360 360
FT CARBOHYD 372 372
FT CARBOHYD 504 504
FT CARBOHYD 515 515
FT CARBOHYD 688 688
FT CARBOHYD 811 811
FT CARBOHYD 816 816
FT CARBOHYD 842 842
FT CARBOHYD 857 857
SQ SEQUENCE 903 AA; 100305 MW; 9742AB1590908AEC CRC64;

alignment_scores: Quality: 248.00 Length: 74

```
RT      mitochondrial functions is identified as IRA2 an attenuator of RAS1
RT      and RAS2 gene products."
RL      Curr Genet. 21:325-329(1992).
CC      -|- FUNCTION: INHIBITORY REGULATOR OF THE RAS-CYCLIC AMP PATHWAY.
CC      -|- SIMILARITY: THE GTPASE ACTIVITY OF RAS PROTEINS.
CC      -|- SIMILARITY: CONTAINS 1 RAS-GAP DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M33779; AAA34710.1; -
DR      EMBL; X83121; CAA58201.1; -
DR      EMBL; X75449; CAA53202.1; -
DR      EMBL; Z74823; CAA99093.1; -
DR      PIR; S11190; RGHY12.
DR      SGD; S0005441; IRA2.
DR      InterPro; IPR001936; RasGAP.
DR      Pfam; PF00616; RasGAP; 1.
DR      SMART; SM00323; RasGAP; 1.
DR      PROSITE; PS00509; RAS_GTPASE_ACTIV_1; 1.
DR      PROSITE; PS50018; RAS_GTPASE_ACTIV_2; 1.
DR      GTPase activation. 1890
KW      DOMAIN 1701
FT      DOMAIN 399
FT      POLY-SER.
FT      DOMAIN 412
FT      POLY-SER.
FT      DOMAIN 520
FT      POLY-ALA.
FT      DOMAIN 2469
FT      POLY-LEU.
FT      CONFLICT 2317
FT      I -> K (IN REF. 3).
SQ      SEQUENCE 3079 AA; 351631 MW; 651EB2A2EBB479C0 CRC64;

alignment_scores:
Quality: 71.50 Length: 58
Ratio: 2.043 Gaps: 3
Percent Similarity: 60.345 Percent Identity: 36.207

alignment_block:
US-09-049-696-5 x IRA2_YEAST
Align seg 1/1 to: IRA2_YEAST from: 1 to: 3079
25 AAACCCACAAAGAGAGCTCCAAACCAAGCAAAATCAAAATGCAATCTC 74
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
190 LysPheAsnThrArgThrLeuGlnIle.....LeuGlnAsnMetIleSe 204
75 CGAAGCACATGGGAGTGCCTGATCCGTGATTCGAGGACTTTAAGAAACAC 124
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
204 rHisValHisGlyAsn.....IleLeuThrThrLeuSerSerI 218
125 TCCTATGACACACAGCCACCAATCCCA.....CCTTCTCAT 162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
218 leLeuProArgHisSerTyThrLeuThrArgHisAsnHisProSerHis 234
163 TGTGTCAGATTGGACAAAGAATTG 186
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
235 CysLysMetIleAspSerThrLeu 242

seq_name: SwissProt_39:APSA_EMENI
seq_documentation_block:
ID APSA_EMENI STANDARD; PRT; 1676 AA.
AC Q00083;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ANUCLEATE PRIMARY STERIGMATA PROTEIN.
GN APSA.
OS Emricella nidulans (Aspergillus nidulans).

seq_documentation_block:
ID IRA2_YEAST STANDARD; PRT; 3079 AA.
AC P19158; Q08239;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE INHIBITORY REGULATOR PROTEIN IRA2.
GN IRA2 OR GLC4 OR CCS1 OR YOL081W OR O0985.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=90318397; PubMed=2164637;
RA Tanaka K., Nakafuku M., Tamaoki F., Kaziro Y., Matsumoto K., Toh-E A.;
RT "IRA2, a second gene of Saccharomyces cerevisiae that encodes a
RT protein with a domain homologous to mammalian ras GTPase-activating
RT protein.";
RL Mol. Cell. Biol. 10:4303-4313(1990).
RN [2]
SEQUENCE OF 1-2423 FROM N.A.
RP STRAIN=S288C / FY1679;
RX MEDLINE=95208358; PubMed=7900427;
RA Zundstein E., Griffin H., Schweizer M.;
RT "Sequence of a 10.27 kb segment on the left arm of chromosome XV from
RT Saccharomyces cerevisiae includes part of the IRA2 gene and a
RT putative new gene.";
RN Yeast 10:1383-1387(1994).
RN [3]
SEQUENCE OF 1982-3079 FROM N.A.
RP MEDLINE=97321807; PubMed=9178509;
RA Tzermia M., Katsoulou C., Alexandraki D.;
RT "Sequence analysis of a 33.2 kb segment from the left arm of yeast
RT chromosome XV reveals eight known genes and ten new open reading
RT frames including homologues of ABC transporters, inositol
RT phosphatases and human expressed sequence tags.";
RL Yeast 13:583-589(1997).
RN [4]
IDENTIFICATION OF CCS1 AS IRA2.
RX MEDLINE=92405229; PubMed=1326414;
RA Bussereau F., Dupont C.H., Boy-Marcotte E., Mallet L., Jacquet M.;
RT "The CCS1 gene from Saccharomyces cerevisiae which is involved in
```


OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 4;
RX MEDLINE=95164553; PubMed=7860626;
RA Fischer R., Timberlake W.E.;
RT "Aspergillus nidulans apsA (nucleate primary sterigmata) encodes a
RT coiled-coil protein required for nuclear positioning and completion
RT of asexual development";
RL J. Cell Biol. 128:485-498(1995).
RN [2]
RN CHARACTERIZATION.
RX MEDLINE=98025472; PubMed=9379904;
RA Aselmann R., Sievers N., Fischer R.;
RT "Nuclear traffic in fungal hyphae: in vivo study of nuclear migration
RT and positioning in Aspergillus nidulans";
RL Mol. Microbiol. 25:757-769(1997).
CC -!- FUNCTION: REQUIRED FOR NUCLEAR POSITIONING AND COMPLETION OF
CC ASEXUAL DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X82289; CAA57733.1; -
DR InterPro: IPR001849; PH.
DR Pfam: PF00169; PH: 1.
DR SMART: SM00233; PH: 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
KW Coiled coil; Membrane.
FT DOMAIN 51 127
FT DOMAIN 193 359
FT DOMAIN 408 453
FT DOMAIN 1393 1504
SQ SEQUENCE 1676 AA; 183671 MW; 8BDAL7F4212AE4D8 CRC64;

alignment_scores:
Quality: 69.50 Length: 51
Ratio: 2.397 Gaps: 2
Percent Similarity: 56.863 Percent Identity: 37.255

alignment_block:
us-09-049-696-5 x APSA_EMENI ..
Align seg 1/1 to: APSA_EMENI from: 1 to: 1676

39 GAAGCTCCAAACAGCAAAATCAAAATGC..... 68
||| ||| ||||| |||||
337 GluAspMetAspThrGlnAsnGlnGluLeuAlaLysAlaValAlaMetAr 353
69AATCCGAGACCATGGAGATGATCCGT.....G 99
353 gLeuArgGlnGlnGluAlaGluSerThrArgGluValAlaValArgProHisA 370
100 ATTCCTGAGGACTTTAAGAAACACCTCTATGACACACACAGCCACCAAAAT 149
||||||| ||||| ||||| ||||| ||||| |||||
370 spSerGluAspGluGlnAlaThrProGluAsnSerProProSer 386
150 CCC 152
|||
387 Pro 387

seq_name: SwissProt_39:CCAF_HUMAN

Seq_documentation_block:
ID CCAF_HUMAN STANDARD; PRT: 1966 AA.
AC O60840; O43901;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VCJTAGE-DEPENDENT L-TYPE CALCIUM CHANNEL ALPHA-1F SUBUNIT.
GN CACNA1F OR CACNA1F.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS CSNB2 D-369; Q-508; W-1049 & H-1364.
RC TISSUE=Retina;
RX MEDLINE=98324776; PubMed=9662399;
RA Strom T.M., Nyakatura G., Apfelstedt-Sylla E., Hellebrand H.,
RA Lorenz B., Weber B.H.F., Wutz K., Gutwilling N., Ruether K.,
RA Drescher B., Sauer C., Zrenner E., Meitinger T., Rosenthal A.,
RA Meindl A.;
RT "An L-type calcium-channel gene mutated in incomplete X-linked
RT congenital stationary night blindness";
RL Nat. Genet. 19:260-263(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Blechschmidt K., Nyakatura G., Strom T.M., Drescher B., Menzel U.,
RA Meindl A., Rosenthal A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE OF 1200-1966 FROM N.A.
RX MEDLINE=98008923; PubMed=9344638;
RA Fisher S.E., Ciccodicola A., Tanaka K., Curci A., Desicato S.,
RA D'Urso M., Craig I.W.;
RT "Sequence-based exon prediction around the synaptophysin locus reveals
RT a gene-rich area containing novel genes in human proximal Xp";
RL Genomics 45:340-347(1997).
CC -!- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
CC ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CC CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CC CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1F
CC GIVES RISE TO L-TYPE CALCIUM CURRENTS. LONG-LASTING (L-TYPE)
CC CALCIUM CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA)
CC GROUP. THEY ARE BLOCKED BY DIHYDROPYRIDINES (DHP),
CC PHENYLALKYLAMINES, BENZOTHAZEPINES, AND BY OMEGA-AGATOXIN-IIIA
CC (OMEGA-AGA-IIIA). THEY ARE HOWEVER INSENSITIVE TO OMEGA-CONOTOXIN-
CC GVIA (OMEGA-CTX-GVIA) AND OMEGA-AGATOXIN-IVA (OMEGA-AGA-IVA).
CC -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: EXPRESSION IN SKELETAL MUSCLE AND RETINA.
CC -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC -!- DISEASE: DEFECTS IN CACNA1F ARE THE CAUSE OF INCOMPLETE X-LINKED
CC CONGENITAL STATIONARY NIGHT BLINDNESS (CSNB2), A NONPROGRESSIVE
CC RETINAL DISORDER CHARACTERIZED BY DECREASED VISUAL ACUITY AND LOSS
CC OF NIGHT VISION.
CC -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

us-09-049-696-5.rsp

Tue Apr 2 09:40:09 2002

entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

```

CC CC EMBL; AJ006216; CRA06916.1; -
CC CC EMBL; AJ224874; CRA12175.1; -
DR DR EMBL; AF235097; AAF62518.1; -
DR DR EMBL; U93305; AAB92359.1; -
DR DR MIM; 300110; -
DR DR MIM; 300071; -
DR DR InterPro; IPR002077; Ca_channel.
DR DR InterPro; IPR002111; Cat_channel_TpL.
DR DR InterPro; IPR000636; Cation_chan_non_lig.
DR DR InterPro; IPR001682; Channel_pore_Ca_Na.
DR DR pfam; PF00520; Ion_trans; 4.
DR DR PRINTS; PR00167; CACHANNEL.
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW Calcium channel; Glycoprotein; Repeat; Multigene family;
KW Calcium-binding; Phosphorylation; Disease mutation; Vision.
KW REPEAT 79 375
FT REPEAT 504 750
FT REPEAT 847 1129
FT REPEAT 1166 1433
FT DOMAIN 1 92
FT DOMAIN 93 111
FT DOMAIN 112 129
FT DOMAIN 130 149
FT DOMAIN 150 161
FT DOMAIN 162 180
FT DOMAIN 181 201
FT DOMAIN 202 220
FT DOMAIN 221 239
FT DOMAIN 240 259
FT DOMAIN 260 347
FT DOMAIN 348 372
FT DOMAIN 373 518
FT DOMAIN 519 538
FT DOMAIN 539 553
FT DOMAIN 554 572
FT DOMAIN 573 580
FT DOMAIN 581 599
FT DOMAIN 600 609
FT DOMAIN 610 628
FT DOMAIN 629 647
FT DOMAIN 648 668
FT DOMAIN 669 722
FT DOMAIN 723 747
FT DOMAIN 748 860
FT DOMAIN 861 879
FT DOMAIN 880 895
FT DOMAIN 896 915
FT DOMAIN 916 927
FT DOMAIN 928 946
FT DOMAIN 947 952
FT DOMAIN 953 972
FT DOMAIN 973 991
FT DOMAIN 992 1011
FT DOMAIN 1012 1101
FT DOMAIN 1102 1122
FT DOMAIN 1123 1179
FT DOMAIN 1180 1198
FT DOMAIN 1199 1213
FT DOMAIN 1214 1233
FT DOMAIN 1234 1262
FT DOMAIN 1263 1279
FT DOMAIN 1280 1299
FT DOMAIN 1300 1318
FT DOMAIN 1319 1338
FT DOMAIN 1339 1405
FT DOMAIN 1406 1430
FT DOMAIN 1431 1966
FT DOMAIN 648 654
FT DOMAIN 783 788

```

```

FT DOMAIN 798 814
FT DOMAIN 1110 1113
FT DOMAIN 1629 1634
FT DOMAIN 395 412
FT SITE 330 330
FT SITE 700 700
FT SITE 1075 1075
FT SITE 1372 1372
FT BINDING 1049 1139
FT BINDING 1386 1452
FT BINDING 1398 1441
FT BINDING 1441 1441
FT MOD_RES 1459 1470
FT CA_BIND 295 295
FT CARBOHYD 369 369
FT VARIANT 508 508
FT VARIANT 1049 1049
FT VARIANT 1364 1364
FT CONFLICT 1225 1225
FT CONFLICT 1265 1271
FT CONFLICT 1652 1767
FT CONFLICT 1849 1849
FT SEQUENCE 1966 AA; 219494 MW; FEB47E19FA57E31D CRC64;

```

alignment_scores: Quality: 69.00 Length: 90
Ratio: 1.438 Gaps: 5
Percent Similarity: 53.333 Percent Identity: 28.889

alignment_block:

US-09-049-696-5/rev x CCAF_HUMAN

Align seg 1/1 to: CCAF_HUMAN from: 1 to: 1966

```

216 TGTTCCTCAGATTTCGACGAGCTAAACACACAATCTTGTCCCAATGTC 167
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
587 CysPheValValCysGlyLeuGluThrLeuValGluValGlu 603
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
166 AGCAATGAGAAGGT...GGGATTTGGTGGTGTCTGTCTAGAGAGTGGT 120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
603 yalaMet.GlnProLeuGlyIleSerValLeuArgCysValArg..... 617
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
119 TTCTTTAAAGTCTCAGATCAGATCAGATCAGATCAGATGCTGCTTCG.... 75
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
618 LeuLeuArgIlePheLysValThrArgHisrPalaSerLeuSerAsnLe 634
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
74.....GAGATTGCAATTTGATTGTCG 53
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
634 uValAlaSerLeuLeuAsnSerMetLysSerIleAlaSerLeuLeuLeu 651
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 TTGTTTGGAGCTTCTTTGTGTGGTGT.....TTGTTTC 21
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
651 LeuLeuPheLeuPheIleIleIlePheSerLeuLeuGlyMetGlnLeuPhe 667
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
20 TGTACGAATTCACATAT 3
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
668 GlyGlyLysPheAsnLeu 673

```

seq_name: SwissProt_39:DPOL_SULAC

seq_documentation_block:
ID: DPOL_SULAC STANDARD; PRT: 875 AA.
AC P95630;

DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA POLYMERASE I (EC 2.7.7.7).
GN DP01 OR POLS.
OS Sulfolobus acidocaldarius.
OC Archaea: Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2285;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;
RX MEDLINE=97080536; PubMed=8921881;
RA Datukishvili N., Pokholok D., Lottspeich F., Prangishvili D.,
Rechinsky V.;
RT "The DNA polymerase-encoding gene from a thermoacidophilic archaeon
Sulfolobus acidocaldarius";
RL Gene 177:271-273(1996).
RN [2]
RP CHARACTERIZATION.
RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;
RX MEDLINE=85269642; PubMed=3927262;
RA Klimczak L.J., Grumet F., Burger K.J.;
RT "Purification and characterization of DNA polymerase from the
archaeobacterium Sulfolobus acidocaldarius";
RL Nucleic Acids Res. 13:5269-5282(1985).
CC -!- FUNCTION: THIS POLYMERASE POSSESSES TWO ENZYMIC ACTIVITIES:
DNA SYNTHESIS (POLYMERASE) AND AN EXONUCLEOLYTIC ACTIVITY THAT
DEGRADES SINGLE STRANDED DNA IN THE 3' TO 5' DIRECTION.
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
N PYROPHOSPHATE + DNA(N).
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U33846; AAC44598.1; -
DR InterPro; IPR002064; DNA_pol_B.
DR Pfam; PF00136; DNA_pol_B; 1.
DR PRINTS; PR00106; DNA_POLB.
DR SMART; SM00486; POLBC; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Hydrolase; Exonuclease.
SQ SEQUENCE 875 AA; 100913 MW; F7E29E5EBCEB8A5 CRC64;

alignment_scores:

Quality:	66.50	Length:	80
Ratio:	1.583	Gaps:	4
Percent Similarity:	52.500	Percent Identity:	27.500

alignment_block:

US-09-049-696-5 x DP01_SULAC

Align seg 1/1 to: DP01_SULAC from: 1 to: 875

3 ATAGTTGAATTCGTACAGACAAACCAACAAAGAACTCCAACAA 52
::: ||||| ::::::::::::::|||
7 LeuPheAspPheSerIleLysLysAsnGluSerLysGluGlnThrAsnG1 23
53 GCAAAAT.....CAAAAATGCAATCTCCGAAGCACATGGG 87
23 nGluSerValGluValProLysGlnThrAlaAsn...ArgThrLysIleG 39
88 AAGTATCGCTGATCTGAGGACTTTAAGAAACCACCTCTATGACAACA 137
|| ||::::::::::|||
39 luTrpIleLysGluAlaGluAspGlyLys..... 48

138 CAGCCACCAATCCACCTCTCATCTGCTCAGATTGGA..... 176
::: |||||
49ValTyPheLeuLeuGlnValAspTyraSpGlyLy 60
177CAAGAATTGTGTTTGTAGTCTCTTGACAAATCTGGA 212
::: ||||| ::::::::::::::|||
60 slySerArgAlaValcysLysLeuTyraSpLysGluGly 73
seq_name: SwissProt_39:CCAS_HUMAN
seq_documentation_block:
ID CCAS_HUMAN STANDARD; PRT; 1873 AA.
AC Q13698; Q13934; Q12896;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VOLTAGE-DEPENDENT L-TYPE CALCIUM CHANNEL ALPHA-1S SUBUNIT (CALCIUM
DE CHANNEL, L TYPE, ALPHA-1 POLYPEPTIDE, ISOFORM 3, SKELETAL MUSCLE).
GN CACNA1S OR CACNLIA3 OR CAC1 OR CACNI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=95229168; PubMed=7713519;
RA Hoggan K., Powers P.A., Gregg R.G.;
RT "Cloning of the human skeletal muscle alpha 1 subunit of the
RT dihydropyridine-sensitive L-type calcium channel (CACNLIA3).";
RL Genomics 24:608-609(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96435439; PubMed=8838325;
RA Hoggan K., Gregg R.G., Powers P.A.;
RT "The structure of the gene encoding the human skeletal muscle alpha 1
RT subunit of the dihydropyridine-sensitive L-type calcium channel
RT (CACNLIA3).";
RL Genomics 31:392-394(1996).
RN [3]
RP SEQUENCE OF 1200-1300 FROM N.A., AND VARIANTS HYPOKPP G-1239 & H-1239.
RX MEDLINE=94273190; PubMed=8004673;
RA Ptacek L.J., Tawil R., Griggs R.C., Engel A.G., Layzer R.B.,
RA Kwiatkowski H., McManis P.G., Santiago L., Moore M., Fouad G.,
RA Bradley P., Leppert M.F.;
RT "Dihydropyridine receptor mutations cause hypokalemic periodic
RT paralysis.";
KL Cell 77:863-868(1994).
RN [45]
RP SEQUENCE OF 1223-1413 FROM N.A.
RA Soldatov N.M.;
RT "Human skeletal muscle L-type Ca2+ channel alpha 1S subunit gene shows
RT splicing patterns similar to alpha 1C and alpha 1D genes in the region
RT involved in hereditary disorders.";
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 788-830; 1019-1085 AND 1293-1318 FROM N.A.
RX MEDLINE=93162636; PubMed=7916735;
RA Gregg R.G., Couch F., Hoggan K., Powers P.A.;
RT "Assignment of the human gene for the alpha-1 subunit of the skeletal
RT muscle DHP-sensitive calcium channel (CACNLIA3) to chromosome 1q31-
RT q32.";
RL Genomics 15:107-112(1993).
RN [6]
RP VARIANT HYPOKPP HIS-528.
RX MEDLINE=95078851; PubMed=7987325;
RA Jurkatt-Rott K., Lehmann-Horn F., Elbaz A., Heine R., Gregg R.G.,
RA Hoggan K., Powers P.A., Lapie P., Vale-Santos J.E., Weissbach J.,
RA Fontaine B.;
RT "A calcium channel mutation causing hypokalemic periodic paralysis.";
RL Hum. Mol. Genet. 3:1415-1419(1994).
RN [7]
RP REVISIONS, VARIANT MSH5 HIS-1086, AND VARIANTS HIS-458 AND CYS-1539.

Align seg 1/1 to: CCAS_HUMAN from: 1 to: 1873

```
216 TGCTTCAGATTTGTCAGGACTAAACACAAATTTCTTCCATGTC 167
||||| :|||:|||||:
501 CyspheValValCysSerGlyIleLeuGluIleLeuValGluSerG1 517
166 AGCAATGAGAAGT...CGGATTGGTGGTGTGTTGTCATGAGAGTGGT 120
||||| :|||:|||||:
517 yAlaMet.ThrProLeuGlyIleSerValLeuArgCysIleArg..... 531
119 TTTCTTAAAGTCCACAAATCAGGATCACTTCCCATGTGCTTGC..... 75
:||||:||||:|||||:|||||:
532 LeuLeuargIlePheIleThrLysTyrTrpThrSerLeuSerAsnLe 548
74 .....GAGATTGCATTTTGTGTTGTC 53
||||| :|||:|||||:
548 uValAlaSerLeuLeuAsnSerIleArgSerIleAlaSerLeuLeuLeu 565
52 TTGTTTGGAGCTTCTTTGTTGTCG 30
||||| :|||:|||||:
565 euLeuPheLeuPheIleValIle 572
```

seq_name: SwissProt_39:RBL2_MOUSE

seq_documentation_block:

ID RBL2_MOUSE STANDARD; PRT: 1135 AA.

AC 064700;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE RETINOBLASTOMA-LIKE PROTEIN 2 (130 KDA RETINOBLASTOMA-ASSOCIATED

PROTEIN) (PRB2) (P130) (RBR-2).

GN RBL2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

SEQUENCE FROM N.A.

RX MEDLINE=96199216; PubMed=8621630;

RA Chen G., Guy C.T., Chen H.W., Hu N., Lee E.Y.H.P., Lee W.H.;

RT "Molecular cloning and developmental expression of mouse p130, a

member of the retinoblastoma gene family.";

RL J. Biol. Chem. 271:9567-9572(1996).

RN [2]

SEQUENCE FROM N.A.

RX MEDLINE=96192345; PubMed=9019172;

RA Pertile P., Baldi A., de Luca A., Virgilio L., Pisano M.M.,

RA Giordano A.;

RT "Molecular cloning, expression, and developmental characterization of

the murine retinoblastoma-related gene Rb2/p130.";

RL Cell Growth Differ. 6:1659-1664(1995).

RN [3]

SEQUENCE FROM N.A.

RX MEDLINE=96203998; PubMed=8622859;

RA Lecouter J.E., Whyte P.F.M., Rudnicki M.A.;

RT "Cloning and expression of the Rb-related mouse p130 mRNA.";

RL Oncogene 12:1433-1440(1996).

CC -!- FUNCTION: MAY HAVE A FUNCTION IN CELL CYCLE REGULATION. BINDS TO

E1A PROTEIN. MAY ACT AS A TUMOR SUPPRESSOR. POTENT INHIBITOR OF

E2F-MEDIATED TRANS-ACTIVATION, ASSOCIATES PREFERENTIALLY WITH

E2F3. BINDS TO CYCLINS A AND E (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC -!- SIMILARITY: BELONGS TO THE RETINOBLASTOMA PROTEIN (RB) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.

```
CC
CC EMBL: U36799; AAB48991.1;
CC EMBL: U50850; AAC52598.1;
CC EMBL: U47333; AAC52555.1;
CC HS3P: P06400; IGUX.
CC MGD: MGI:105085; Rbl2.
CC InterPro: IPR000553; Cyclin.
CC InterPro: IPR002720; RB_A.
CC InterPro: IPR002719; RB_B.
CC Pfam: PF01857; RB_A; 1.
CC Pfam: PF01858; RB_B; 1.
CC SMART: SM00385; CYCLIN; 2.
CC Transcription regulation; DNA-binding; Nuclear protein; Cell cycle;
CC Phosphorylation; Anti-oncogene.
```

```
FT DOMAIN 414 1021 POCKET (BINDS E1A).
FT DOMAIN 414 613 DOMAIN A.
FT DOMAIN 614 824 SPACER.
FT DOMAIN 825 1021 DOMAIN B.
FT DOMAIN 9 13 POLY-PRO.
FT DOMAIN 14 17 POLY-ALA.
FT DOMAIN 20 24 POLY-GLU.
FT DOMAIN 992 995 POLY-GLU.
FT CONFLICT 211 253 MISSING (IN REF. 3).
```

```
FT CONFLICT 341 341 A -> P (IN REF. 2).
FT CONFLICT 342 342 A -> P (IN REF. 3).
FT CONFLICT 381 381 R -> T (IN REF. 1).
FT CONFLICT 428 428 T -> S (IN REF. 1).
FT CONFLICT 431 431 S -> T (IN REF. 1).
FT CONFLICT 443 443 Q -> R (IN REF. 2 AND 3).
FT CONFLICT 484 493 MISSING (IN REF. 2).
FT CONFLICT 768 768 A -> R (IN REF. 1).
FT CONFLICT 826 826 P -> A (IN REF. 3).
```

```
FT CONFLICT 947 948 PT -> RA (IN REF. 3).
FT CONFLICT 1023 1023 S -> A (IN REF. 2 AND 3).
FT CONFLICT 1044 1044 V -> I (IN REF. 1).
SQ SEQUENCE 1135 AA; 127472 MW; 0565B4F998ACCAOD CRC64;
```

alignment_scores:

Quality: 63.50 Length: 72
Ratio: 1.628 Gaps: 4
Percent Similarity: 54.167 Percent Identity: 29.167

alignment_block:

US-09-049-696-5 x RBL2_MOUSE

Align seg 1/1 to: RBL2_MOUSE from: 1 to: 1135

```
17 TACAGACAAACACACAAAGAAC..... 43
|||||:|||||:|||||:|||||
912 TyrArgThrGlnProGlnAlaArgSerGlnValTyrArgSerValLeu11 928
41 .....TCCAAACAAGCAAAATCAAAATGC 68
928 eLysGlyLysArgArgAsnSerGlySerSerGlySerArgSerHisGlnA 945
69 AATCTCCGAAGCACATGGAGTAGTCGATTCTGAGGAC..... 110
|||||:|||||:|||||:|||||
945 snSerProThrGluLeuAsnThrAsp.ArgAlaSerArgAspSerSerPr 961
111 .....TTTAAAGAAACCACTCTCTATG.....ACAACACAGC 141
961 oValMetArgSerAsnSerThrLeuProValProGlnProSerSerAlap 978
142 CACCAAAATCCCAAC 155
978 roProThrProThr 982
```

seq_name: SwissProt_39:O2B2_HUMAN

seq_documentation_block:

ID O2B2_HUMAN STANDARD; PRT: 357 AA.

Q9CZK3; Q9GZL2; Q9VZ99;
20-AUG-2001 (Rel. 40, Created)
20-AUG-2001 (Rel. 40, Last sequence update)
20-AUG-2001 (Rel. 40, Last annotation update)
OLFACTORY RECEPTOR 2B2 (OLFACTORY RECEPTOR 6-1) (OR6-1) (HS6M1-10).
OR2B2.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_taxid=9606;
[1]
SEQUENCE FROM N.A., AND VARIANT ARG-234.
Ziegler A., Ehlers A., Forbes S.A., Trowsdale J., Uchanska-Ziegler B.,
Volz A., Younger R., Beck S.;
"Polymorphic olfactory receptor genes and HLA loci constitute extended
haplotypes";
(In) Kasahara M. (eds.);
Major histocompatibility complex-evolution, structure, and function,
pp.110-130, Springer-Verlag, Tokyo (2000).
[2]
SEQUENCE OF 1-310 FROM N.A.
Wild A.;
Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
-1- FUNCTION: PUTATIVE ODORANT RECEPTOR.
-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute, there are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AJ302584; CAC20504.1; -
EMBL; AJ302585; CAC20505.1; -
DR DR EMBL; AJ302586; CAC20506.1; -
DR DR EMBL; AJ302587; CAC20507.1; -
DR DR EMBL; AJ302588; CAC20508.1; -
DR DR EMBL; AJ302589; CAC20509.1; -
DR DR EMBL; AJ302590; CAC20510.1; -
DR DR EMBL; AJ302591; CAC20511.1; -
DR DR EMBL; AJ302592; CAC20512.1; -
DR DR EMBL; AJ302593; CAC20513.1; -
DR DR EMBL; 298744; CABL1427.1; -
InterPro: IPR000276; GPCR_Rhodopsn.
RefSeq: PF00001; 7tm_1; 1.
PRINTS; PR00237; GPCRRHODOPSIN.
DR DR PROSITE; PS00237; G-PROTEIN_RECEP_FL1; 1.
DR DR PROSITE; PS00262; G-PROTEIN_RECEP_FL2; 1.
DR DR G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction; Polymorphism.
KW DOMAIN 1; 25
FT DOMAIN 1 25
FT TRANSMEM 26 49
FT DOMAIN 50 57
FT TRANSMEM 58 79
FT DOMAIN 80 100
FT TRANSMEM 101 120
FT DOMAIN 121 139
FT TRANSMEM 140 158
FT DOMAIN 159 195
FT TRANSMEM 196 219
FT DOMAIN 220 236
FT TRANSMEM 237 259
FT DOMAIN 260 272
FT TRANSMEM 273 292
FT DOMAIN 293 357
FT DISULFID 97 189
FT CARBOHYD 5 5
FT VARIANT 234
SEQUENCE 357 AA; 40412 MW; 53E2F828208C440A CRC64;
/FtId=VAR_010943.

```

alignment_scores:      63.00      Length:      91
                        Quality:      1.340      Gaps:      5
                        Ratio:      51.648      Percent Identity: 28.571
Percent Similarity: 51.648

alignment_block:
US-09_049-696-5/rev x O2B2_HUMAN

Align, Seg 1/1 to: O2B2_HUMAN from: 1 to: 357

216 CTTCCAGATTGTCAAGGACTTAACACACAATT.....CTTGTCC 174
    ||| ||| ||| ||| ||| ||| ||| ||| |||
68 LeuLeuAspLeuCysThrThrSerThrValProGlnMetLeuValAs 84

173 APTCTCAGCAATAGAGAGTG...GGATTGGTGGCTGTGTTCATAG 127
    ||| ||| ||| ||| ||| ||| ||| ||| |||
84 nllCysAsnThrArgLysValIleSerThrGlyGlyCysValAlaGlnL 101

126 GAGTGGTTTCTTAAAG.....TCTCTCAGATCAGGATCATTCC... 86
    ||| ||| ||| ||| ||| ||| ||| ||| |||
101 euPheIlePheLeuAlaLeuGlySerThrGlucCysLeuLeuAlaVal 117

85 .....CATGCTCTCGGAGATTGCATTTT..... 62
    ||| ||| ||| ||| ||| ||| ||| ||| |||
118 MetCysPheAspArgPheValAlaIleCysArgProLeuHisTyrSerI 134

61 .....TGATTTCCTGTTTGGAGGCTCTTTGTTGGTGGTTT 25
    ||| ||| ||| ||| ||| ||| ||| ||| |||
134 elleMetHisGlnArgLeuCysPheGlnLeuAlaAlaSerTrpIleS 151

24 GTTCTGTACAGATTCAACTATA 2
    ||| ||| ||| ||| ||| ||| ||| ||| |||
151 ergLysPheSerAsnSerValLeu 158

seq_name: SwissProt_39:KAD_PARDE

seq_documentation_block:
ID: KAD_PARDE STANDARD: PRT; 217 AA.
AC: P10772; P77922;
DT: 01-JUL-1989 (Rel. 11, Created)
DT: 15-JUL-1998 (Rel. 36, Last sequence update)
DT: 20-AUG-2001 (Rel. 40, Last annotation update)
DE: ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP TRANSPHOSPHORYLASE).
GN: ADXN.
OS: Paracoccus denitrificans.
OC: Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC: Paracoccus.
OX: NCST_taxID=266;
RN: [1]
FP: SEQUENCE FROM N.A.
FC: STRAIN=7A.11T;
RA: MEDLINE=99077072; PubMed=9862211;
RA: Perrier V., Burlacu-Miron S., Boussac A., Meier A., Gilles A.M.;
RT: "Metal chelating properties of adenylate kinase from Paracoccus
RT: denitrificans.";
RL: Protein Eng. 11:917-923(1998).
RN: [2]
SEQUENCE
MEDLINE=89153098; PubMed=2537726;
FX: Spurgin P., Tomasselli A.G., Schiltz E.;
RA: "The amino acid sequence of adenylate kinase from Paracoccus
RT: denitrificans and its relationship to mitochondrial and microbial
RT: adenylate kinases.";
RL: Eur. J. Biochem. 179:621-628(1989).
CC: -1- FUNCTION: THIS SMALL UBIQUITOUS ENZYME IS ESSENTIAL FOR
CC: -1- MAINTENANCE AND CELL GROWTH.
CC: -1- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC: -1- SUBUNIT: MONOMER.
CC: -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC: -1- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.

```


Ratio: 1.824 Gaps: 0
Percent Similarity: 60.714 Percent Identity: 26.786

alignment_block:

US-09-049-696-5 x COMM_DROME

Align seg 1/1 to: COMM_DROME from: 1 to: 370

```
27 AACCAACAAGAGCTCCCAACAAAGCAAAATCAAAATGCAATCTCCG 76
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
266 AsnTrpGluGlnGluProLeuAsnLysLeuGlnLysCysLysLeuSe 282
77 AAGCAGATGGGAGTGATCGGTGATCTTCAGGACTTTAAGAAACCACTC 126
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
282 rAlaAlaAlaValGluGluAspLysAlaAspSerSerSerThr 299
127 CTATGACACACAGCAGCAACCAATCCACCTTCTGTCAGATTGGA 176
|||||:|||||:|||||:|||||:|||||:|||||:
299 erAlaSerAlaSerProSerSerSerGluSerSerAsnLeuAlaThra 315
177 CAAGAAGTTGTGTTTA 194
|||||:
316 ThrProAlaIleCysIle 321
```

seq_name: SwissProt_39:SUV3_YEAST

seq_documentation_block:

AC P32580; 002649; PRT; 737 AA:
DT 01-OCT-1993 (Rel. 25, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MITOCHONDRIAL ATP-DEPENDENT RNA HELICASE SUV3 PRECURSOR.
GN SUV3 OR YPL029W OR LPB2W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;

SEQUENCE FROM N.A.

AC P32580; 002649; PRT; 737 AA:
DT 01-OCT-1993 (Rel. 25, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MITOCHONDRIAL ATP-DEPENDENT RNA HELICASE SUV3 PRECURSOR.
GN SUV3 OR YPL029W OR LPB2W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;

AC P32580; 002649; PRT; 737 AA:
DT 01-OCT-1993 (Rel. 25, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MITOCHONDRIAL ATP-DEPENDENT RNA HELICASE SUV3 PRECURSOR.
GN SUV3 OR YPL029W OR LPB2W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;

AC P32580; 002649; PRT; 737 AA:
DT 01-OCT-1993 (Rel. 25, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MITOCHONDRIAL ATP-DEPENDENT RNA HELICASE SUV3 PRECURSOR.
GN SUV3 OR YPL029W OR LPB2W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;

AC P32580; 002649; PRT; 737 AA:
DT 01-OCT-1993 (Rel. 25, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MITOCHONDRIAL ATP-DEPENDENT RNA HELICASE SUV3 PRECURSOR.
GN SUV3 OR YPL029W OR LPB2W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;

DR Pfam: PF00271; helicase_C; 1.
KW SMART; SM00490; HELIC; 1.
KW Mitochondrion; Transit peptide; Helicase; ATP-binding; RNA-binding.
FT TRANSIT 1 25 MITOCHONDRIAL ATP-DEPENDENT RNA HELICASE
FT CHAIN 26 737 SUV3
FT N-BIND 239 246 ATP (POTENTIAL).
FT VARIANT 272 272 V -> L (IN SUV3-1; PROBABLY EXCISED
INTRONS ARE LESS EFFICIENTLY RELEASED).
FT CONFLICT 165 165 W -> R (IN REF. 1).
FT CONFLICT 664 664 E -> D (IN REF. 1).
SQ SEQUENCE 737 AA; 84324 MW; 17D18200E5D18099 CRC64;

alignment_scores:

Quality: 62.00 Length: 48
Ratio: 1.938 Gaps: 1
Percent Similarity: 66.667 Percent Identity: 33.333

alignment_block:

US-09-049-696-5 x SUV3_YEAST

Align seg 1/1 to: SUV3_YEAST from: 1 to: 737

```
63 AAATGCAATCTCCGAAGCATGCGTGTCTGAGGACTT 112
|||||:|||||:|||||:|||||:|||||:|||||:
282 ArgCysAsnLeuLeuThrGlyGluValIleArgAspLeuAspPar 298
113 TAAGAAACCACTCTATGACACACAGCCACCAATCCCACTCTCAT 162
|||||:|||||:|||||:|||||:|||||:|||||:
298 gGlyAsnSerAlaGlyLeuThr.....SerGlyThrValGlu 311
163 TGCTGCAGATTGGCAAGAATTGTGTGTTAGTCCTTGACAAA 206
|||||:|||||:|||||:|||||:|||||:|||||:
311 etValProIleAsnGlnLysPheAspValValValLeuAspGlu 325
```

seq_name: SwissProt_39:CCAS_RABIT

seq_documentation_block:

ID CCAS_RABIT STANDARD; PRT; 1873 AA.
AC P07293;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE VOLTAGE-DEPENDENT L-TYPE CALCIUM CHANNEL ALPHA-1S SUBUNIT (CALCIUM CHANNEL, L TYPE, ALPHA-1 POLYPEPTIDE, ISOFORM 3, SKELETAL MUSCLE).
GN CACNAIS OR CACNL1A3 OR CAC1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Skeletal muscle;
RX MEDLINE=87258269; PubMed=3037387;
RA Tababe T., Takeshima H., Mikami A., Flockerzi V., Takahashi H.,
Kajigawa K., Kojima M., Matsuo H., Hirose T., Numa S.;
RT "Primary structure of the receptor for calcium channel blockers from skeletal muscle.";
RL Nature 328:313-318(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=88336904; PubMed=2458626;
RA Ellis S.B., Williams M.E., Ways N.R., Brenner R., Sharp A.H.,
Leung A.T., Campbell K.P., McKenna E., Koch W.J., Hui A., Schwartz A.,
Harpold M.M.;
RT "Sequence and expression of mRNAs encoding the alpha 1 and alpha 2 subunits of a DHP-sensitive calcium channel.";
RL Science 241:1661-1664(1988).
RN [3]
RP BETA-SUBUNIT BINDING DOMAIN.
RX MEDLINE=94150724; PubMed=7509046;

RA Pragnell M., de Waard M., Mori Y., Tanabe T., Snutch T.P.,
RT Campbell K.P.;
RT "Calcium channel beta-subunit binds to a conserved motif in the I-II
RL cytoplasmic linker of the alpha 1-subunit.";
RN Nature 368:67-70(1994).
RX [4]
RP PHENYLALANINE-BINDING SITE.
RX MEDLINE-91067656; PubMed-2174553;
RA Striessnig J., Glossmann H., Catterall W.A.;
RT "Identification of a phenylalanine binding region within the alpha 1
RL subunit of skeletal muscle Ca2+ channels.";
RN Proc. Natl. Acad. Sci. U.S.A. 87:9108-9112(1990).
RX [5]
RP DIHYDROPYRIDINE-BINDING SITE.
RX MEDLINE-92021019; PubMed-1656465;
RA Nakayama Y., Taki M., Striessnig J., Glossmann H., Catterall W.A.,
RT Kanakawa Y.;
RT "Identification of 1,4-dihydropyridine binding regions within the
RL alpha 1 subunit of skeletal muscle Ca2+ channels by photoaffinity
labeling with diazepam.";
RN Proc. Natl. Acad. Sci. U.S.A. 88:9203-9207(1991).
RX [6]
RP DIHYDROPYRIDINE-BINDING SITE.
RX MEDLINE-92073369; PubMed-1660150;
RA Striessnig J., Murphy B.J., Catterall W.A.;
RT "Dihydropyridine receptor of L-type Ca2+ channels: identification of
RL binding domains for [3H](+)-PN200-110 and [3H]azidopine within the
alpha 1 subunit.";
RN Proc. Natl. Acad. Sci. U.S.A. 88:10769-10773(1991).
RX [7]
RP PHOSPHORYLATION OF SER-687 AND SER-1617.
RX MEDLINE-89008428; PubMed-2844809;
RA Roehrkasten A., Meyer H.E., Nastainczyk W., Sieber M., Hofmann F.;
RT "CAMP-dependent protein kinase rapidly phosphorylates serine-687 of
RL the skeletal muscle receptor for calcium channel blockers.";
RN J. Biol. Chem. 263:15325-15329(1988).
RX [8]
RP PHOSPHORYLATION BY CAPK.
RX MEDLINE-89367340; PubMed-2549550;
RA Nunoki K., Florio V., Catterall W.A.;
RT "Activation of purified calcium channels by stoichiometric protein
RL phosphorylation.";
RN Proc. Natl. Acad. Sci. U.S.A. 86:6816-6820(1989).
CC -!- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1S
GIVES RISE TO L-TYPE CALCIUM CURRENTS. LONG-LASTING (L-TYPE)
CALCIUM CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA)
GROUP. THEY ARE BLOCKED BY DIHYDROPYRIDINES (DHP).
PHENYLALANINES, BENZOTHAZEPINES, AND BY OMEGA-AGATOXIN-IIIa
(OMEGA-AGA-IIIA). THEY ARE HOWEVER INSENSITIVE TO OMEGA-CONOTOXIN-
GVIA (OMEGA-CTX-GVIA) AND OMEGA-AGATOXIN-IVA (OMEGA-AGA-IVA).
CALCIUM CHANNELS CONTAINING THE ALPHA-1S SUBUNIT PLAY AN IMPORTANT
ROLE IN EXCITATION-CONTRACTION COUPLING IN SKELETAL MUSCLE.
-!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY. AN
ADDITIONAL GAMMA SUBUNIT IS PRESENT ONLY IN SKELETAL MUSCLE L-TYPE
CHANNEL.
-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
-!- TISSUE SPECIFICITY: SKELETAL MUSCLE SPECIFIC.
-!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
-!- DOMAIN: THE LOOP BETWEEN REPEATS II AND III INTERACTS WITH THE

RYANODINE RECEPTOR, AND IS THEREFORE IMPORTANT FOR CALCIUM RELEASE
FROM THE ENDOPLASMIC RETICULUM NECESSARY FOR MUSCLE CONTRACTION.
-!- PTM: THE ALPHA-1S SUBUNIT IS FOUND IN TWO ISOFORMS IN THE SKELETAL
MUSCLE: A MINOR FORM OF 212 KDA CONTAINING THE COMPLETE AMINO ACID
SEQUENCE, AND A MAJOR FORM OF 190 KDA DERIVED FROM THE FULL-LENGTH
FORM BY POST-TRANSLATIONAL PROTEOLYSIS CLOSE TO PHE-1690.
-!- PTM: BOTH THE MINOR AND MAJOR FORMS ARE PHOSPHORYLATED IN VITRO BY
CAPK. PHOSPHORYLATION BY CAPK STIMULATES THE CALCIUM CHANNEL
FUNCTION!
-!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
FAMILY.
-!- FUNCTION:
-!- This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; X05921; CAA29355.1; -;
EMBL; M23919; AAA31159.1; -;
PIR; A30063; A30063.
DR InterPro: IPR002077; Ca_Channel.
DR InterPro: IPR002111; Cat_channel_TrpL.
DR InterPro: IPR000636; Cation_chan_non_lig.
DR InterPro: IPR001682; Channel_pore_Ca_Na.
DR Pfam: PF00520; ion_trans; 4.
DR PRINTS: PR00167; CACHANNEL.
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW Calcium channel; Glycoprotein; Repeat; Multigene family;
KW Calcium-binding; Phosphorylation; Polymorphism.
FT REPEAT 38 337 I.
FT REPEAT 418 664 II.
FT REPEAT 786 1068 III.
FT REPEAT 1105 1384 IV.
FT DOMAIN 1 51 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 52 70 S1 OF REPEAT I (POTENTIAL).
FT DOMAIN 71 88 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 89 108 S2 OF REPEAT I (POTENTIAL).
FT DOMAIN 109 120 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 121 139 S3 OF REPEAT I (POTENTIAL).
FT DOMAIN 140 160 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 161 179 S4 OF REPEAT I (POTENTIAL).
FT DOMAIN 180 198 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 199 218 S5 OF REPEAT I (POTENTIAL).
FT DOMAIN 219 309 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 310 334 S6 OF REPEAT I (POTENTIAL).
FT DOMAIN 335 432 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 433 451 S1 OF REPEAT II (POTENTIAL).
FT DOMAIN 452 466 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 467 486 S2 OF REPEAT II (POTENTIAL).
FT DOMAIN 487 494 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 495 513 S3 OF REPEAT II (POTENTIAL).
FT DOMAIN 514 523 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 524 542 S4 OF REPEAT II (POTENTIAL).
FT DOMAIN 543 561 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 562 581 S5 OF REPEAT II (POTENTIAL).
FT DOMAIN 582 636 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 637 661 S6 OF REPEAT II (POTENTIAL).
FT DOMAIN 662 799 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 800 818 S1 OF REPEAT III (POTENTIAL).
FT DOMAIN 819 834 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 835 854 S2 OF REPEAT III (POTENTIAL).
FT DOMAIN 855 866 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 867 885 S3 OF REPEAT III (POTENTIAL).
FT DOMAIN 886 892 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 893 911 S4 OF REPEAT III (POTENTIAL).
FT DOMAIN 912 930 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 931 950 S5 OF REPEAT III (POTENTIAL).
FT DOMAIN 951 1040 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1041 1065 S6 OF REPEAT III (POTENTIAL).
FT DOMAIN 1066 1118 CYTOPLASMIC (POTENTIAL).

[illegible]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

DR EMBL; AL022727; CA018782.1; -
DR EMBL; AJ302537; CAC20462.1; -
DR EMBL; AJ302538; CAC20463.1; -
DR EMBL; AJ302539; CAC20464.1; -
DR EMBL; AJ302540; CAC20465.1; -
DR EMBL; AJ302541; CAC20466.1; -
DR EMBL; AJ302542; CAC20467.1; -
DR EMBL; AJ302543; CAC20468.1; -
DR EMBL; AJ302544; CAC20469.1; -
DR EMBL; AJ302545; CAC20470.1; -
DR EMBL; AJ302546; CAC20471.1; -
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PROSITE; PS00237; G-PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS0262; G-PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction.
FT DOMAIN 1 25 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 26 49 1 (POTENTIAL).
FT DOMAIN 50 57 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 58 79 2 (POTENTIAL).
FT DOMAIN 80 100 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 101 120 3 (POTENTIAL).
FT DOMAIN 121 139 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 140 158 4 (POTENTIAL).
FT DOMAIN 159 195 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 196 219 5 (POTENTIAL).
FT DOMAIN 220 236 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 237 259 6 (POTENTIAL).
FT DOMAIN 260 272 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 273 292 7 (POTENTIAL).
FT DOMAIN 293 313 CYTOPLASMIC (POTENTIAL).
FT DISULFID 97 189 BY SIMILARITY.
FT CARBOHYD 5 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 313 AA; 35542 MW; 786BF1EE7B624440 CRC64;

alignment_scores:

Quality: 61.00 Length: 72
Ratio: 1.564 Gaps: 3
Percent Similarity: 54.167 Percent Identity: 25.000

alignment_block:

US-09-049-696-5/rev x 02B3_HUMAN ..

Align seg 1/1 to: 02B3_HUMAN from: 1 to: 313

169 TGCAGCAATGAGAAGTGGATGCTGGTGTGTCATAGAGTGGT 120
|||||.....
87 CysAsnLysThrIleSerYrAlaGlyCysValAlaHisLeuIleI 103
119 TTTCTTAAAG.....TCCTCAGAAATCAGGATCACTTCC..... 86
|||||.....
103 ePheLeuAlaLeuGlyAlaThrGlyCysLeuLeuAlaValMetSerP 120
85CATGTCCTTCGAGATTCGAT..... 65
120 heAspArgTyrValAlaValAlaCysArgProLeuHisTyrValIleMet 136
64TTTGTGTTTGTGTTTGTGAGGCTTCTTGTGCTGTTTGTCTGT 18
....|||||.....
137 AsnTyrTrpPheCysLeuArgMetAlaAlaPheSerTrpLeuIleGlyPh 153
17 ACAGAAATCAACTATA 2
|||||.....

I55 eGlyAsnSerValLeu 158

seq_name: SwissProt_39:Y147_HAEIN

seq_documentation_block:

ID Y147_HAEIN STANDARD; PRT; 633 AA.

AC P43543;

DT 01 NOV-1995 (Rel. 32, Created)

DT 20 AUG-2001 (Rel. 32, Last sequence update)

DE HYPOTHETICAL PROTEIN HI0147.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

QC Haemophilus.

GX NCBI_TaxID=727;

RN SEQUENCE FROM N.A.

RC STRAIN-RD / KW20 / ATCC 51907;

EX MEHLIN-95350630; PubMed-7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kevlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,

RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,

RA Kiper L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,

RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,

RA Venter J.C.;

"Whole-genome random sequencing and assembly of Haemophilus

Influenzae Rd."

Science 289:496-512(1995).

CC SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE

CC (PROBABLE).

CC SIMILARITY: BELONGS TO THE YIAN/YGIK FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)

EMBL; U32700; AAC21819.1; -

DR TIGR; HI0147; -

DR InterPro: IPR000252; DedA.

DR Pfam: PF00597; DedA; 1.

KW Hypothetical protein; Transmembrane; Inner membrane;

KW Complete proteome.

FT TRANSMEM 26 46 POTENTIAL.

FT TRANSMEM 53 73 POTENTIAL.

FT TRANSMEM 100 120 POTENTIAL.

FT TRANSMEM 134 154 POTENTIAL.

FT TRANSMEM 170 190 POTENTIAL.

FT TRANSMEM 213 233 POTENTIAL.

FT TRANSMEM 261 281 POTENTIAL.

FT TRANSMEM 305 325 POTENTIAL.

FT TRANSMEM 349 369 POTENTIAL.

FT TRANSMEM 374 394 POTENTIAL.

FT TRANSMEM 424 444 POTENTIAL.

FT TRANSMEM 448 468 POTENTIAL.

FT TRANSMEM 476 496 POTENTIAL.

FT TRANSMEM 522 542 POTENTIAL.

FT TRANSMEM 544 564 POTENTIAL.

FT TRANSMEM 569 589 POTENTIAL.

FT TRANSMEM 604 624 POTENTIAL.

SQ SEQUENCE 633 AA; 69378 MW; B96F7D97286C06DA CRC64;

alignment_scores:

Quality: 61.00 Length: 61
Ratio: 1.488 Gaps: 5

[illegible]

A;Residues: 1-1676 <FIS>

A;Molecule type: DNA

A;Molecule type: DNA
A;Residues: 1-2308, 'V', 2310-2316, 'I', 2318-3079 <TAN>

7 LeupheAspPheSerIleLysLysAsnGluSerLysGluGlnThrAsnGlu 23

Ты не раскисай, душенька, не плачь, не плачь.

119 AspLeuCysArgLysPheAsnSerAsnLysGluGlnProTyrProAspAs 135

```

alignment_scores:      Length:      75
                        Quality:    64.00
                        Ratio:     1.455
                        Gaps:       4
                        Percent Similarity: 32.000

alignment_block:
alignment_596-5/rev.y.255645

```


216 TGCTTCCAGATTGCTCAAGGACTAAACACACAATCTTTGTGCCAATCGC 167
 501 CysPheValValCysSerGlyIleLeuGluLeuLeuValGluSerG1 517
 166 AGCAATGAGAAGGT...GGAGTTTGGTGGCTGTGTGTACATAGGAGTGGT 120
 517 yAlaMet.ThrProLeuGlyIleSerValLeuArgCysIleArg..... 531
 119 TTTCTTAAAGTCCTCAAGATCCAGGATCCATTCCTCCATGTGCTGTG.... 75
 532 LeuLeuArgLeuPheLysIleThrLysTyrTrpThrSerLeuSerAsnLe 548
 74GACATTCGATCTTTTGATTTCG 53
 548 uValAlaSerLeuLeuAsnSerIleArgSerIleAlaSerLeuLeuLeuL 565
 52 TTCTTTGGAGCTTCTTTGTTGTG 30
 565 euLeuPheLeuPheIleIleIle 572

This Page Blank (uspto)

OM of: US-09-049-696-5 to: Issued_Patents_AA.* out_format : pfs

Date: Mar 30, 2002 2:32 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL-frame+ n2p.model -DEV-xlp
-Q/cgn2_1/USPTO.spool/US09049696/runat_28032002_145238_2012/app_query.fasta_1.12579
-DGAP-issued_Patents_AA -QFMT-fastan -SUFFIX=rai -GAPOP=12.000
-BAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-OGAPOP=6.000 -OGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=DIsum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696@cgn1_109 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-049-696-5
Query length: 220
Database: Issued_Patents_AA.*
Database sequences: 212232
Database length: 2250322
Search time (sec): 362.110000
```

score_list:	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/2/iaa/PCRTUS_COMB.ppt:US94-00198-4	+	71.50	128.79	0.3286	3079
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-145-006C-12	+	63.00	129.63	2.54	357
/cgn2_6/ptodata/2/iaa/PCRTUS_COMB.ppt:US94-00545-12	+	63.00	129.63	2.54	357
/cgn2_6/ptodata/2/iaa/backfiles1.ppt:5386025-6	-	62.00	110.91	5.35	1872
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-435-675B-4	-	62.00	110.91	5.35	1873
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-336-257A-7	-	62.00	110.91	5.35	1873
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-07-745-206A-2	-	62.00	109.50	5.56	2161
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-455-543A-49	-	62.00	109.50	5.56	2161
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-455-543A-51	-	62.00	109.50	5.56	2161
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-223-305C-49	-	62.00	109.50	5.56	2161
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-223-305C-51	-	62.00	109.50	5.56	2161
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-311-363-2	-	62.00	109.50	5.56	2161
/cgn2_6/ptodata/2/iaa/PCRTUS_COMB.ppt:US94-07297-39	+	61.00	118.22	5.61	699
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-396-479B-2	+	61.00	115.50	6.04	921
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-818-823-2	+	60.50	112.72	7.35	1082
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-106-493A-2	+	60.50	112.72	7.35	1082
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-429-264-2	+	60.50	112.21	7.45	1139
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-832-883-2	+	60.50	112.21	7.45	1139
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-832-877-2	+	60.50	111.59	7.57	1213
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-937-102-2	+	60.50	100.09	10.30	3898
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-059-853-2	-	60.50	100.09	10.30	3898
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-07-745-206A-7	-	60.00	105.62	10.03	1968
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-455-543A-45	-	60.00	105.62	10.03	1968
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-223-305C-45	-	60.00	105.62	10.03	1968
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-311-363-7	+	59.00	110.97	11.09	897
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-515-884-6	+	59.00	110.97	11.09	897
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-019-095A-18	+	58.50	131.68	7.19	97
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-019-095A-8	+	58.50	115.37	11.13	508
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-019-095A-22	+	58.50	115.12	11.21	521
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-019-095A-2	+	58.50	115.03	11.23	526
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-145-006C-5	+	58.00	108.65	15.05	890
/cgn2_6/ptodata/2/iaa/PCRTUS_COMB.ppt:US94-00545-5	+	58.00	108.65	15.05	890
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-08-896-449A-2	-	57.50	96.11	23.77	2813
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-132-652-2	-	57.50	96.11	23.77	2813
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-111-470-2	-	57.00	119.28	14.44	237
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-149-934-3	+	57.00	112.35	17.38	479
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-07-762-132A-2	-	57.00	109.83	18.60	619
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-301-722A-4	-	57.00	109.83	18.60	619

```
/cgn2_6/ptodata/2/iaa/backfiles1.ppt:5206163-3 - 57.00 95.63 27.20 2616 1 p
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-614-801A-6 - 56.50 113.54 19.02 376
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-338-530A-2 - 56.00 110.90 23.05 435
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-267-384-2 - 56.00 110.90 23.05 435 1
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-091-432-2 - 56.00 109.14 24.16 520 1
seq_name: /cgn2_6/ptodata/2/iaa/PCRTUS_COMB.ppt:US94-00198-4
```

```
seq_documentation_block:
; Sequence 4, Application PC/TUS9400198
; GENERAL INFORMATION:
; APPLICANT: Schering Corp.
; TITLE OF INVENTION: RAS Associated GAP Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Schering Corp.
; STREET: 1 Girald Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: 6.0.8
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00198
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/004,824
; FILING DATE: 15-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lunn, Paul G.
; REGISTRATION NUMBER: 32,743
; REFERENCE/DOCKET NUMBER: DX0352 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)822-7255
; TELEFAX: (201)822-7039
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3079 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; PCT-US94-00198-4
```

```
alignment_scores:
Quality: 71.50 Length: 58
Ratio: 2.043 Gaps: 3
Percent Similarity: 60.345 Percent Identity: 36.207
```

```
alignment_block:
```

```
US-09-049-696-5 x PCT-US94-00198-4 ..
```

```
Align seg 1/1 to: PCT-US94-00198-4 from: 1 to: 3079
```

```
25 AAACCAACAAAGAGCTCCAAACAGCAAAATCAAAATCAATCTC 74
||| :||||| :||||| :||||| :|||||
190 LysPheAsnThrArgThrLeuGlnIle.....LeuGlnAsnMetIleSe 204
75 CGAAGCACATCGGAAGTGATCCGTGATTCTGAGAGCTTTAAGAAACAC 124
||||| :||||| :||||| :||||| :|||||
204 rHisValHisGlyAsn.....IleLeuThrThrLeuSerSerI 218
125 TCCTATGACAACACACGCCCAATCCCA.....CCTCTCAT 162
|||| :|||| :|||| :|||| :||||
218 IleuProArgHisLysSerTyLeuThrArgHisAsnHisProSerHis 234
```

us-09-049-696-5.ra1

Tue Apr 2 09:40:08 2002

```
163 TGCTGACGATTGGACAAAGATTG 186
||||:||||:|
235 CysLysMetIleAspSerThrLeu 242
||||:||||:|

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-145-006C-12

seq_documentation_block:
; Sequence 12, Application US/08145006C
; Patent No. 5656452
; GENERAL INFORMATION:
; APPLICANT: Rao, Anjana
; APPLICANT: Hogan, Patrick Gerald
; APPLICANT: McCaffrey, Patricia
; APPLICANT: Jain, Jugnu
; TITLE OF INVENTION: NF-ATp, A T LYMPHOCYTE
; TITLE OF INVENTION: DNA-BINDING PROTEIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/145.006C
; FILING DATE: October 29, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/017,052
; FILING DATE: February 11, 1993
; APPLICATION NUMBER: 08/006,067
; FILING DATE: January 15, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 04590/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-145-006C-12

alignment_scores:
Quality: 63.00 Length: 59
Ratio: 1.969 Gaps: 0
Percent Similarity: 54.237 Percent Identity: 30.508

alignment_block:
US-09-049-696-5 x US-08-145-006C-12 ..
Align seg 1/1 to: US-08-145-006C-12 from: 1 to: 357

14 CTGTACAGAACAAACCAAGAGCTCCAAACAGCAAAATCAAA 63
||||:||||:|
264 MetTrpLysThrSerProAspSerProValSerArgAlaProSerLy 280
||||:||||:|
64 ATGCAATCTCCAGCACCATGGGAAGTGCCTGGATTCGAGGACATT 113
||||:||||:|
280 saLaGlyLeuProArgHisIleTyProAlaValGluPheLeuGlyProC 297
||||:||||:|
```

us-09-049-696-5.ra1

```
114 AAGAAACCACTCTATGACACACACACCAACCAATCCACCTTCTCATT 163
||||:||||:|
297 ysGluGlnGlyGluArgAsnSerAlaProGluSerIleLeuVal 313
||||:||||:|

164 GCTGCAGATTGGACAAAGATTGTGTG 190
||||:||||:|
314 ProProThrTrpProLysProLeuVal 322
||||:||||:|

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US94-00545-12

seq_documentation_block:
; Sequence 12, Application PC/TUS9400545
; GENERAL INFORMATION:
; APPLICANT: Rao, Anjana
; APPLICANT: Hogan, Patrick Gerald
; APPLICANT: McCaffrey, Patricia
; APPLICANT: Jain, Jugnu
; TITLE OF INVENTION: NF-ATp, A T LYMPHOCYTE
; TITLE OF INVENTION: DNA-BINDING PROTEIN
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00545
; FILING DATE: 18-JAN-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/145,006
; FILING DATE: October 29, 1993
; APPLICATION NUMBER: 08/017,052
; FILING DATE: February 11, 1993
; APPLICATION NUMBER: 08/006,067
; FILING DATE: January 15, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 04590/007001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; PCT-US94-00545-12

alignment_scores:
Quality: 63.00 Length: 59
Ratio: 1.969 Gaps: 0
Percent Similarity: 54.237 Percent Identity: 30.508

alignment_block:
US-09-049-696-5 x PCT-US94-00545-12 ..
Align seg 1/1 to: PCT-US94-00545-12 from: 1 to: 357

14 CTGTACAGAACAAACCAAGAGCTCCAAACAGCAAAATCAAA 63
||||:||||:|
```

264 MetTrpIysThrSerProAspProSerProValSerArgAlaProSerLy 280
64 AATGCAATCTCCGAAGACACATGGAGTGTGCTGATCTGAGGACATTT 113
113:|||||
280 sAlaGlyLeuProArgHisIleTyrProAlaValGluPheLeuGlyProC 297
114 AAGAAACCACTCTATGACACACAGCCACCAAAATCCACCTTCATTT 163
114:|||||
297 ySGLuGlnGlyGluArgAsnSerAlaProGluSerIleLeuLeuVal 313
164 GCTGCAGATTGGACAAAGATTGTGTG 190
114:|||||
314 ProThrTrpProLysProLeuVal 322
seq_name: /cgn2_6/ptodata/2/1aa/backfiles1.pep:5386025-6
seq_documentation_block:
; Patent No. 5386025
; APPLICANT: JAY, SCOTT D.; ELLIS, STEVEN B.; HARPOOLD, MICHAEL
; M.; CAMPBELL, KEVIN P.
; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/482,384
; FILING DATE: 20-FEB-1990
; SEQ ID NO: 6:
; LENGTH: 1872
5386025-6

alignment_scores:
Quality: 62.00 Length: 75
Ratio: 1.409 Gaps: 4
Percent Similarity: 58.667 Percent Identity: 30.667

alignment_block:

US-09-049-696-5/rev x 5386025-6 ..

Align seg 1/1 to: 5386025-6 from: 1 to: 1872

216 TGCTTCCAGATTGTGCAAGGACTAAACACACAATCTTTGTCCAATCTGC 167
||||| :|||||
501 CysPheValValCysSerGlyIleLeuLeuLeuValGluSerGI 517
166 AGCAATGAGAAGGT...GGGATTGGTGGTGTGTTCTCATAGAGTGGT 120
:||||| :|||||
517 yAlaMet.ThrProLeuGlyIleSerValLeuArgCysIleArg..... 531
119 TTTCTTAAAGTCTCAGATCAGGATCACTTCCCATGTCTTCG..... 75
:||||| :|||||
532 LeuLeuArgLeuPheIleThrLysTyrTrpThrSerLeuSerAsnLe 548
74GAGATTGCATTTTGTGTTGC 53
548 uValAlaSerLeuLeuAsnSerIleArgSerIleAlaSerLeuLeuLe 565
52 TTGTTGGAGCTTCTTTGTGTG 30
:||||| :|||||
565 euLeuPheLeuPheIleIle 572

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-435-675B-4

seq_documentation_block:

; Sequence 4, Application US/08435675B

; Patent No. 5710250

; GENERAL INFORMATION:

; APPLICANT: Ellis, Steven Bradley

; APPLICANT: Williams, Mark E.

; APPLICANT: Harpold, Michael Miller

; APPLICANT: Schwartz, Arnold

; APPLICANT: Brenner, Robert

; TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS

; NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Brown, Martin, Haller & McClain

STREET: 1660 Union Street

CITY: San Diego

STATE: CA

COUNTRY: USA

ZIP: 92101-2926

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/435,675B

FILING DATE: 05-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/314,083

FILING DATE: 28-SEP-1994

APPLICATION NUMBER: US 07/914,231

FILING DATE: 13-JUL-1992

APPLICATION NUMBER: US 07/603,751

FILING DATE: 08-NOV-1990

ATTORNEY/AGENT INFORMATION:

NAME: Seigman, Stephanie L

REGISTRATION NUMBER: 33,779

REFERENCE/DOCKET NUMBER: 6362-53193

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-238-0999

TELEFAX: 619-238-0062

TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1873 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

US-08-435-675B-4

alignment_scores:

Quality: 62.00 Length: 75

Ratio: 1.409 Gaps: 4

Percent Similarity: 58.667 Percent Identity: 30.667

alignment_block:

US-09-049-696-5/rev x US-08-435-675B-4 ..

Align seg 1/1 to: US-08-435-675B-4 from: 1 to: 1873

216 TGCTTCCAGATTGTGCAAGGACTAAACACACAATCTTTGTCCAATCTGC 167
||||| :|||||
501 CysPheValValCysSerGlyIleLeuLeuLeuValGluSerGI 517
166 AGCAATGAGAAGGT...GGGATTGGTGGTGTGTTCTCATAGAGTGGT 120
:||||| :|||||
517 yAlaMet.ThrProLeuGlyIleSerValLeuArgCysIleArg..... 531
119 TTTCTTAAAGTCTCAGATCAGGATCACTTCCCATGTCTTCG..... 75
:||||| :|||||
532 LeuLeuArgLeuPheIleThrLysTyrTrpThrSerLeuSerAsnLe 548
74GAGATTGCATTTTGTGTTGC 53
548 uValAlaSerLeuLeuAsnSerIleArgSerIleAlaSerLeuLeuLe 565
52 TTGTTGGAGCTTCTTTGTGTG 30
:||||| :|||||
565 euLeuPheLeuPheIleIle 572

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-336-257A-7

```

|||||: |||||:|||||:
555 eureuxPhLeuphellelrie 572

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-745-206A-2

seq_documentation_block:
; Sequence 2, Application US/07745206A
; Patent No. 5429921
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark
; APPLICANT: McCue, Ann
; APPLICANT: Feldman, Daniel
; TITLE OF INVENTION: Human Calcium Channel Compositions and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitch, Even, Tabin & Flannery
; STREET: 135 S. Lasalle
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/745,206A
; FILING DATE: 19910815
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Feder, Scott B
; REFERENCE/DOCKET NUMBER: 51504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-372-7842
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2161 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-745-206A-2

```

[illegible]

52 TTGTTTGGAGCTTCTTTGTTGGTT.....TTGTTTC 21
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
656 euleupheleupheillellepheSerLeuLeuGlyMetGlnLeuPhe 672
20 TGTACAGAATTCAACTAT 3
:|||||:|||||:
673 GlyGlyLysPheAsnPhe 678

seq_name: /cgn2_6/ptodata/2/laa/5A_COMB.pep:US-08-455-543A-49

seq_documentation_block:
; Sequence 49, Application US/08455543A

; Patent No. 5792846

; GENERAL INFORMATION:

; APPLICANT: Harpold, Michael

; APPLICANT: Ellis, Steven

; APPLICANT: Williams, Mark

; APPLICANT: Feldman, Daniel

; APPLICANT: McCue, Ann

; APPLICANT: Brenner, Robert

; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND

; METHOD OF INVENTION: METHODS

; NUMBER OF SEQUENCES: 57

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Brown, Martin, Haller & McClain

; STREET: 1660 Union Street

; CITY: San Diego

; STATE: California

; COUNTRY: USA

; ZIP: 92101-2926

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ Version 1.5

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/455,543A

; FILING DATE: May 31, 1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/223,305

; FILING DATE: April 4, 1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/868,354

; FILING DATE: April 10, 1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/745,206

; FILING DATE: 15-AUG-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/620,250

; FILING DATE: 30-NOV-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/482,384

; FILING DATE: 20-FEB-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/603,751

; FILING DATE: 04-APR-1989

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/US89/01408

; FILING DATE: 04-APR-1989

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/176,899

; FILING DATE: 04-APR-1988

; ATTORNEY/AGENT INFORMATION:

; NAME: Seidman, Stephanie L.

; REGISTRATION NUMBER: 33,779

; REFERENCE/DOCKET NUMBER: 6362-52517

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619)238-0999

; TELEFAX: (619)238-0062

; INFORMATION FOR SEQ ID NO: 49:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2161 amino acids

; TYPE: amino acid

;

;

;

;

;

;

;

;

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-455-543A-49

Alignment_scores:
; Quality: 62.00 Length: 90
; Ratio: 1.292 Gaps: 5
; Percent Similarity: 53.333 Percent Identity: 26.667

Alignment_block:

US-09-049-696-5/rev x US-08-455-543A-49 ..

Align_seg 1/1 to: US-08-455-543A-49 from: 1 to: 2161

216 TGCTTCAGATTGTCAGGACTAAACACACAATCTTTGTCACAATCTGC 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
592 CysPheValValCysGlyGlyIleThrGluThrIleLeuValGluLeu 608
166 AGCAATGAGAAAGT...GGGATTGGTGGCTGTCTTCATAGGAGTGGT 120
|||:|||||:|||||:|||||:|||||:|||||:|||||:
608 uIleMet.SerProLeuGlyIleSerValPheArgCysValArg..... 622
119 TTTCTTAAAGTCTCAGATCAGGATCCTCCATGTGCTTCG..... 75
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
623 LeuLeuArgIlePheLysValThrArgHisTrpThrSerLeuSerAsnLe 639
7.....GAGATTGCATTTTTCATTTTGC 53
639 uValAlaSerLeuLeuAsnSerMetLysSerIleAlaSerLeuLeu 656
52 TTGTTGGAGCTTCTTTGTTGGTT.....TTGTTTC 21
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
656 euleupheleupheillellepheSerLeuLeuGlyMetGlnLeuPhe 672
20 TGTACAGAATTCAACTAT 3
:|||||:|||||:
673 GlyGlyLysPheAsnPhe 678

seq_name: /cgn2_6/ptodata/2/laa/5A_COMB.pep:US-08-455-543A-51

seq_documentation_block:

; Sequence 51, Application US/08455543A

; Patent No. 5792846

; GENERAL INFORMATION:

; APPLICANT: Harpold, Michael

; APPLICANT: Ellis, Steven

; APPLICANT: Williams, Mark

; APPLICANT: Feldman, Daniel

; APPLICANT: McCue, Ann

; APPLICANT: Brenner, Robert

; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND

; METHOD OF INVENTION: METHODS

; NUMBER OF SEQUENCES: 57

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Brown, Martin, Haller & McClain

; STREET: 1660 Union Street

; CITY: San Diego

; STATE: California

; COUNTRY: USA

; ZIP: 92101-2926

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ Version 1.5

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/455,543A

; FILING DATE: May 31, 1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/223,305

;

;

;

;

;

;

;

```

; FILING DATE: April 4, 1994
; PRIOR APPLICATION DATA: 07/868,354
; FILING DATE: April 10, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/745,206
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/620,250
; FILING DATE: 30-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/482,384
; FILING DATE: 20-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,751
; FILING DATE: 04-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US89/01408
; FILING DATE: 04-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-52517
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)238-0999
; TELEFAX: (619)238-0062
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2161 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-455-543A-51

alignment_scores:
  Quality: 62.00      Length: 90
  Ratio: 1.292       Gaps: 5
  Percent Similarity: 53.333   Percent Identity: 26.667

alignment_block:
US-09-049-696-5/rev x US-08-455-543A-51
Align seg 1/1 to: US-08-455-543A-51 from: 1 to: 2161
216 TGCTTCAGATTGTCAGAGGACTAAACACACAAATCTTTGTCCAATCTGC 167
592 CysPheValValCysGlyGlyLeuThrLeuValGluLeuG1 608
166 AGCAATGAGAAGT...GGGATTTGGTGGCTGTGTGCATAGGAGTGGT 120
608 u1eMet.SerProLeuGlyIleSerValPheArgCysValArg..... 622
119 TTTCTTAAGTCTCAGAAATCAGGATCAGTCCCATGTGTTGTC..... 75
623 LeuLeuArgIlePheLeuValThrArgHisItrpThrSerLeuSerAsnLe 639
74 .....GAGATTCATTTTGGTTTC 53
639 uValAlaSerLeuLeuAsnSerMetLysSerIleAlaSerLeuLeuL 656
52 TTGTTTGGAGCTCTTTGTTGTTGTT.....TTGTTTC 21
656 euLeuPheLeuPheIleIleIleIlePheSerLeuLeuGlyMetClnLeuPhe 672
20 TGTACAGAAATCACTAT 3
673 GlyGlyLysPheAsnPhe 678

seq_name: /cgn2_5/ptodata/2/iaa/5B_COMB.pep.US-08-223-305C-49
seq_documentation_block:
; Sequence 49, Application US/08223305C
; Patent No. 5851824
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark
; APPLICANT: Feldman, Daniel
; APPLICANT: McCue, Ann
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/223,305C
; FILING DATE: April 4, 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/868,354
; FILING DATE: April 10, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/745,206
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/620,250
; FILING DATE: 30-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/482,384
; FILING DATE: 20-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,751
; FILING DATE: 04-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US89/01408
; FILING DATE: 04-APR-1988
; APPLICATION NUMBER: US 07/176,899
; FILING DATE: 04-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 52516 (P519739)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)238-0999
; TELEFAX: (619)238-0062
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2161 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-223-305C-49

alignment_scores:
  Quality: 62.00      Length: 90
  Ratio: 1.292       Gaps: 5
```

FILING DATE: 20-FEB-1990

```

; TITLE OF INVENTION: Human Calcium Channel Compositions and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,363
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/745,206
; FILING DATE: 15-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-51506
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)238-0999
; TELEFAX: (619)238-0062
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-311-363-2

alignment_scores:
  Quality: 62.00 Length: 90
  Ratio: 1.292 Gaps: 5
  Percent Similarity: 53.333 Percent Identity: 26.667

alignment_block:
US-09-049-696-5/rev x US-08-311-363-2 ...
Align seg 1/1 to: US-08-311-363-2 from: 1 to: 2161

216 TGCTCCAGATTGTCAGGACTAAACACACAATCTTTGCCAATCTGC 167
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
592 CysPheValValCysGlyGlyIleThrGluThrIleLeuValGluLeu 608
166 AGCAATGAGAAGT...CGGATTGGTGGCTGTGTGTCATAGAGTGGT 120
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
608 uIleMet.SerProLeuGlyIleSerValPheArgCysValArg..... 622
119 TTTCTTAAAGTCCTCAGAAATCAGGATCATCTCCATGCTGCTCG..... 75
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
623 LeuLeuArgIlePheLeuValThrArgHisThrPheSerLeuSerAsnLe 639
74 .....GAGATTGCATTTTGTGATTTTGC 53
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
639 uValAlaSerLeuLeuAsnSerMetLysSerIleAlaSerLeuLeuLeu 656
52 TTGTTTGGAGCTTCTTTCTTGTGGTT.....TTGTTTC 21
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
656 euLeuPheLeuPheIleIleIlePheSerLeuLeuGlyMetGlnLeuPhe 672
20 TGTACAGAAATCAACTAT 3
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
673 GlyGlyLysPheAsnPhe 678

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US94-07297-39
```

```

seq_documentation_block:
; Sequence 39, Application PC/TUS9407297
; GENERAL INFORMATION:
; APPLICANT: Arai, Naoko
; APPLICANT: Masuda, Esteban S.
; APPLICANT: Tokumitsu, Hiroshi
; TITLE OF INVENTION: PURIFIED COMPONENTS OF MAMMALIAN
; TITLE OF INVENTION: TRANSCRIPTION REGULATION COMPLEXES, AND ANALOGS
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John H. C. Blasdale, Schering-Plough Corporation, M-3-W
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh Iicx
; OPERATING SYSTEM: System Software 7.1
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07297
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/222,626
; FILING DATE: 04-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,061
; FILING DATE: 05-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/113,971
; FILING DATE: 30-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,998
; FILING DATE: 30-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/088,483
; FILING DATE: 06-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0392K4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 699 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-07297-39

alignment_scores:
  Quality: 61.00 Length: 59
  Ratio: 1.968 Gaps: 0
  Percent Similarity: 52.542 Percent Identity: 30.508

alignment_block:
US-09-049-696-5 x PCT-US94-07297-39 ..
Align seg 1/1 to: PCT-US94-07297-39 from: 1 to: 699

14 CTGTACAGACAAACCAACCAACAAAGAGCTCCAAACAGCAAAATCAAA 63
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
103 MetTrpLysThrSerProAspProSerProValSerAlaAlaProSerLy 119
64 AATGCAATCTCCGAGACATGGAAGTGCCTGATTCGAGGACTTT 113
:||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
119 salaglyLeuProArgHisIleTyrProAlaValGluPheLeuGlyProc 136
```

355 ysGluGlnGlyGluArqArqAAsnSerAlaProGluSerIleLeuLeuVal 371

```

14  CTGTAGAACAAACACCAACAAGAAAGCTCCAAACAAGAAATCAAA 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
32 MetTlpLysThrSerPronspProSerProValSerAlaAlaProSerLy 338
   : : : : : : : : : : : : : : : : : : : : : : : : : :
64  AATGCAATCTCCGAAGCACATCGGAAGTAGCTCGATTCTCAGAGTTT 113
   : : : : : : : : : : : : : : : : : : : : : : : : : :
338  salaGlyLeuProArgHisIleTyrProAlaValGluPheLeuGlyProC 355
   : : : : : : : : : : : : : : : : : : : : : : : : : :
114  AGAAACACCACTCTATGACACACACAGCCACCAATCCACACTTCTCAT 163
   : : : : : : : : : : : : : : : : : : : : : : : : : :
355  ysGluGlnGlyGluArgAsnSerAlaProGluSerIleLeuLeuVal 371
   : : : : : : : : : : : : : : : : : : : : : : : : : :
164  GCTGCAGATTGGACAAAGAATTGTGTG 190
   : : : : : : : : : : : : : : : : : : : : : : : : : :

```

372 ProProThrTrpProLysProLeuVal 380

seq_name: /SIDS2/gcgcdata/geneseq/geneseqp/AA2001.DAT: AAG75614

seq_name: /SIDS2/gcgcdata/geneseq/geneseqp/AA2001.DAT: AAG75614

US
Homo sapiens.
XX

PN WO200122920-A2.

XX
XX

PD 05-APR-2001.

XX

PF 28-SEP-2000; 2000WO-US26524.

[illegible]

PR 29-SEP-1999; 990S-015/137.
 03-NOV-2000; 000S-0152200

03-NOV-1999; 9905-0103280.
XX

PA (HÜMA-) HÜMAN GENOME SCI INC.

XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
WPI: 2001-235357/24.
XX N-PSDB: AAH35019.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 11: Page 7851-7854; 9803pp: English.
XX
XX AAH32943 to AAH37195 and AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 869 AA;
XX
alignment_scores:
Quality: 386.00 Length: 72
Ratio: 5.361 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-049-696-5 x AAG75614

```

alignment_scores:
  Quality: 386.00
  Ratio: 5.361
  Percent Similarity: 100.000
  Length: 72
  Gaps: 0
  Percent Identity: 100.000

alignment_block:
  US-09-049-696-5 x AAG75614

```

Align seg 1/1 to: AAG75614 from: 1 to: 869

```

3 ATAGTTGAATTTCTGACAGAACAAACCAACAAAGAGCTCCAAACAA 52
|||||
201 lleValGluPheCysThrGluGlnAsnHisAsnLysGluAlaProAsnLy 217
|||||
53 GCAAAATCAAAATGCAATCTCCGAAGCAGCATGGGAAGTGCCTGATT 102
|||||
217 sGlnAsnGlnLysCysAsnLeuArgSerThrTrpGluValIleArgAspS 234
|||||
103 CTGAGGACTTTAAGAAACCACTCTATGACACACACACACCAAAATCCC 152
|||||
234 erGluAspPheLysThrThrProMetThrThrGlnProProAsnPro 250
|||||
153 ACCTTCTCATTTCTGCAGATTGGACAAAGAAATTTGTCTTTAGTCTTTGA 202
|||||
251 ThrPheSerLeuLeuGlnIleGlyGlnArgIleValCysLeuValLeuAs 267
|||||
203 CAATCTGGAAGCATG 218
|||||
267 pLysSerGlySerMet 272

```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74824

seq_documentation_block:

ID AAB74824 standard; Protein: 914 AA.

XX AC AAB74824;

XX DT 13-JUN-2001 (first entry)

XX DE Human ICACC-1 protein sequence.

XX KW ICACC-2; asthma; atopic allergy; asthma-related disorder;
interleukin 9 induced calcium activated chloride channel; IL-9;
calcium activated chloride channel; anti-allergic; anti-asthmatic;
anti-inflammatory; immunomodulatory; cystic fibrosis;
inflammatory bowel disease; autoimmune disease.

XX OS Homo sapiens.

XX PN WO9944620-A1.

XX PD 10-SEP-1999.

XX PF 03-MAR-1999; 99WO-US04703.

XX PR 03-MAR-1998; 98US-0076815.

XX PA (MAGA-) MAGATININ PHARM INC.

XX PI Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McEane M;

XX PI Nicolaides NC, Zhou Y, Dong Q;

XX WPI: 1999-550979/46.

DR N-PSDB: AAF81927.

XX New nucleic acid encoding calcium activated chloride channel, used to

XX identify, e.g. specific modulators for treating atopic allergy -

XX Claim 11; Fig 4B; 75pp; English.

XX The present sequence represents the human interleukin 9 (IL-9) induced
calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
have anti-allergic, anti-asthmatic, anti-inflammatory and
immunomodulatory activities. Compounds (A) that downregulate ICACC are
used to alleviate asthma (or more generally atopic allergy), while those
(B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
inflammatory bowel disease (IBD) (or other autoimmune diseases).
Measuring levels of ICACC-1 can be used in the diagnosis of asthma
(increased levels) or IBD (reduced levels), also for monitoring
treatment of these conditions. The ICACC proteins can be used:

CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
CC identify modulators and binding partners. ICACC polynucleotides can be
CC used to generate transgenic animals or recombinant cells, used to screen
CC for antagonists, also as a source of therapeutic antisense agents or
CC diagnostic probes (for quantifying mRNA expression, e.g. for
CC identification of modulators).

XX Sequence 914 AA;

alignment_scores:
Quality: 386.00 Length: 72
Ratio: 5.361 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-5 x AAB74824

Align seg 1/1 to: AAB74824 from: 1 to: 914

```

3 ATAGTTGAATTTCTGACAGAACAAACCAACAAAGAGCTCCAAACAA 52
|||||
246 lleValGluPheCysThrGluGlnAsnHisAsnLysGluAlaProAsnLy 262
|||||
53 GCAAAATCAAAATGCAATCTCCGAAGCAGCATGGGAAGTGCCTGATT 102
|||||
262 sGlnAsnGlnLysCysAsnLeuArgSerThrTrpGluValIleArgAspS 279
|||||
103 CTGAGGACTTTAAGAAACCACTCTATGACACACACACCAAAATCCC 152
|||||
279 erGluAspPheLysThrThrProMetThrThrGlnProProAsnPro 295
|||||
153 ACCTTCTCATTTCTGCAGATTGGACAAAGAAATTTGTCTTTAGTCTTTGA 202
|||||
296 ThrPheSerLeuLeuGlnIleGlyGlnArgIleValCysLeuValLeuAs 312
|||||
203 CAATCTGGAAGCATG 218
|||||
312 pLysSerGlySerMet 317

```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAM24514

seq_documentation_block:

ID AAM24514 standard; Protein: 914 AA.

XX AC AAM24514;

XX DT 12-OCT-2001 (first entry)

XX DE C902P predicted amino acid sequence.

XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;
immunogenic; gene therapy; vaccine; colonic cancer.

XX OS Homo sapiens.

XX PN WO200149716-A2.

XX PD 12-JUL-2001.

XX PF 29-DEC-2000; 2000WO-US35596.

XX PR 30-DEC-1999; 99US-0476296.

XX PR 10-JAN-2000; 2000US-0480321.

XX PR 15-FEB-2000; 2000US-0504629.

XX PR 06-MAR-2000; 2000US-0519444.

XX PR 19-MAY-2000; 2000US-0575251.

XX PR 29-JUN-2000; 2000US-0609448.

XX PR 28-AUG-2000; 2000US-0649811.

XX (CORI-) CORIXA CORP.

153 ACCTTCTCATTGCTGCAGATTGGACAAAGAAATGTGTGTTAGTCCTTGA 202
|||||
296 ThrPheSerLeuLeuGlnIleGlyGlnArgIleValCysLeuValLeuAs 312
|||||
203 CAAATCTGGAAGCATG 218
|||||
312 PylsSerGlySerMet 317

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB74733

seq_documentation_block:

ID AAB74733 standard; Protein; 914 AA.

XX AAB74733;

DT 12-JUN-2001 (first entry)

DE Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.

XX Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW immunostimulant; cytostatic; cardiac; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
KW antialzheimer; antiparkinsonian; antimicrobial; vulnery; gene therapy;
KW immune disorder; hyperproliferative disorder; cardiovascular disease;
KW cancer; angiogenic disorder; neurological disorder; infectious disease;
KW wound healing; regeneration; chemotaxis; chromosome 1.

OS Homo sapiens.

XX WO200112775-A2.

XX 22-FEB-2001.

XX 16-AUG-2000; 2000WO-US22325.

XX 17-AUG-1999; 99US-0149182.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ni J, Florence KA, Fiscella M, Wei P, Baker KP;
PI Birse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;

XX WPI; 2001-147550/15.

DR N-PSDB; AAF81787.

XX Nucleic acids encoding 25 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -

XX Claim 11; Page 459-460; 485pp; English.

XX AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733
CC to AAB74772. Human secreted proteins can have activities based on the
CC tissues and cells they are expressed in. Example of activities include:
CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;
CC antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiac;
CC vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic;
CC anticonvulsant; antialzheimers; antiparkinsonian; antimicrobial; and
CC vulnery. Human secreted proteins can be used in gene therapy and
CC vaccine. Human secreted protein nucleotide sequences (NAM1) and proteins
CC (PEP1) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. For example, NAM1
CC and PEP1 may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patients genome
CC that affect the activity of proteins by expressing inactive proteins or
CC to supplement the patients own production of polypeptides. Disorders that
CC may be prevented, diagnosed and/or treated include immune disorders,
CC hyperproliferative disorders (e.g. cancers), cardiovascular diseases,
CC angiogenic disorders, neurological disorders, infectious diseases and/or
CC for promoting wound healing, regeneration and /or chemotaxis. AAF81778 to
CC AAF81786 and AAB74732 represent sequences used in the exemplification of

CC the present invention.

XX Sequence 914 AA;

alignment_scores:

Quality: 386.00 Length: 72
Ratio: 5.361 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-5 x AAB74733 ..

Align seg 1/1 to: AAB74733 from: 1 to: 914

3 ATAGTTGAATTTCTACAGACAAACCAACCAACCAAGAGCTCCAAACAA 52
|||||
246 IleValGluPheCysThrGluGlnAsnHisAsnLysGluAlaProAsnLy 262
|||||
53 GCAAAATCAAAAATGCAATCTCCGAGACACATGGGAAGTCGTGATT 102
|||||
262 sGlnAsnGlnLysCysAsnLeuArgSerThrTrpGluValIleArgAsps 279
|||||
103 CTGAGGACTTTTAAAGAAACCACTCTCTATGACACACACACCAAAATCCC 152
|||||
279 erGluAspPheLysLysThrThrProMetThrThrGlnProProAsnPro 295
|||||
153 ACCTTCTCATTGCTGCAGATTGGACAAAGAAATGTGTGTTAGTCCTTGA 202
|||||
296 ThrPheSerLeuLeuGlnIleGlyGlnArgIleValCysLeuValLeuAs 312
203 CAAATCTGGAAGCATG 218
|||||
312 PylsSerGlySerMet 317

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAG75474

seq_documentation_block:

ID AAG75474 standard; Protein; 925 AA.

XX AAG75474;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:6238.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; chromosome 1.

OS Homo sapiens.

XX WO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

XX 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI; 2001-235357/24.

XX N-PSDB; AAB74879.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -

XX Claim 11; Page 7686-7690; 9803pp; English.

XX AAB72943 to AAB73195 and AAG77788 represent human colon


```
seq_documentation_block:
ID   AAB73715 standard; Protein; 913 AA.
XX
AC   AAB73715;
XX
DT   11-SEP-2001 (first entry)
XX
DE   Mouse Gob-5 protein, SEQ ID NO:1.
XX
KW   Mouse Gob-5; murine; goblet cell; human CLCA1 orthologue; drug screening;
KW   expression inhibition; antisense therapy; gene therapy; bronchial asthma;
KW   chronic obstructive pulmonary disease; antiasthmatic.
XX
OS   Mus sp.
XX
PN   WO200138530-A1.
XX
PD   31-MAY-2001.
XX
PF   22-NOV-2000; 2000WO-JP08232.
XX
PR   24-NOV-1999; 99JP-0333479.
PR   27-APR-2000; 2000JP-0127589.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Nakanishi A, Morita S;
XX
WPI: 2001-355935/37.
DR   N-PSDB; AAH46101, AAH46120.
XX
PT   New antisense nucleotide, useful for treatment and prevention of
PT   bronchial asthma and chronic obstructive pulmonary disease -
XX
PS   Claim 1; Page 72-76; 104pp; Japanese.
XX
CC   The invention relates to an antisense nucleotide targetted to the mouse
CC   Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
CC   the CLCA1 gene (coding sequence shown in AAH46102). The invention also
CC   relates to an antibody specific for the Gob-5 protein, medical and
CC   diagnostic compositions containing the antisense nucleotide or the
CC   antibody, and methods and kits for screening for compounds which inhibit
CC   the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.
CC   The antisense oligonucleotides and antibody are therefore useful for the
CC   treatment and prevention of bronchial asthma and chronic obstructive
CC   pulmonary disease. The present sequence represents mouse Gob-5 protein.
XX
SQ   Sequence 913 AA;

alignment_scores:
Quality: 343.00      Length: 72
Ratio: 4.900        Gaps: 0
Percent Similarity: 97.222 Percent Identity: 86.111

alignment_block:
US-09-049-696-5 x AAB73715
..
Align seg 1/1 to: AAB73715 from: 1 to: 913

3 ATAGTTGAATTCGTACAGACAAACCAACAAAGAGCTCCAAACAA 52
.....
247 ValValGluPheCysThrGluLysAsnHisAsnGlnGluAlaProAsnAs 263
.....
53 GCAAAATCAAAATGCAATCTCGAAGCACATGGGAAGTGCCTGGATT 102
.....
263 pGlnAsnGlnArgCysAsnLeuArgSerThrTrpGluValIleGlnGln 280
.....
103 CTGAGGACITTAGAAACACACTCTCTATGACACACAGCCACAAATCCC 152
.....
280 erGluAspPheLysGlnThrThrProMetThrAlaGlnProAlaPro 296
.....
153 ACCTTCTCATTCGTGCAGATTGCAGAACAAAGAAATGTGTGTAGTCTCTTGA 202
.....
```

```
seq_documentation_block:
ID   AAY66749 standard; protein; 919 AA.
XX
AC   AAY66749;
XX
DT   05-APR-2000 (first entry)
XX
DE   Membrane-bound protein PRO1124.
XX
KW   Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW   pharmaceutical; receptor immunoadhesin; gene mapping.
XX
OS   Homo sapiens.
XX
PN   WO9963088-A2.
XX
PD   09-DEC-1999.
XX
PF   02-JUN-1999; 99WO-US12252.
XX
PR   02-JUN-1998; 98US-0087607.
PR   02-JUN-1998; 98US-0087609.
PR   02-JUN-1998; 98US-0087759.
PR   03-JUN-1998; 98US-0087827.
PR   04-JUN-1998; 98US-0088021.
PR   04-JUN-1998; 98US-0088025.
PR   04-JUN-1998; 98US-0088028.
PR   04-JUN-1998; 98US-0088029.
PR   04-JUN-1998; 98US-0088030.
PR   04-JUN-1998; 98US-0088033.
PR   04-JUN-1998; 98US-0088326.
PR   05-JUN-1998; 98US-0088167.
PR   05-JUN-1998; 98US-0088202.
PR   05-JUN-1998; 98US-0088212.
PR   05-JUN-1998; 98US-0088217.
PR   05-JUN-1998; 98US-0088655.
PR   05-JUN-1998; 98US-0088722.
PR   05-JUN-1998; 98US-0088730.
PR   05-JUN-1998; 98US-0088734.
PR   05-JUN-1998; 98US-0088738.
PR   05-JUN-1998; 98US-0088740.
PR   05-JUN-1998; 98US-0088741.
PR   05-JUN-1998; 98US-0088742.
PR   05-JUN-1998; 98US-0088810.
PR   05-JUN-1998; 98US-0088811.
PR   05-JUN-1998; 98US-0088824.
PR   05-JUN-1998; 98US-0088825.
PR   05-JUN-1998; 98US-0088826.
PR   05-JUN-1998; 98US-0088858.
PR   05-JUN-1998; 98US-0088861.
PR   05-JUN-1998; 98US-0088863.
PR   05-JUN-1998; 98US-0088876.
PR   05-JUN-1998; 98US-0089090.
PR   05-JUN-1998; 98US-0089105.
PR   05-JUN-1998; 98US-0089440.
PR   05-JUN-1998; 98US-0089512.
PR   05-JUN-1998; 98US-0089514.
PR   05-JUN-1998; 98US-0089532.
PR   05-JUN-1998; 98US-0089538.
PR   05-JUN-1998; 98US-0089598.
PR   05-JUN-1998; 98US-0089599.
PR   05-JUN-1998; 98US-0089600.
PR   05-JUN-1998; 98US-0089653.
PR   05-JUN-1998; 98US-0089801.
PR   05-JUN-1998; 98US-0089801.
```

PR	16-AUG-1998	98US-0096959
PR	18-AUG-1998	98US-0096960
PR	18-AUG-1998	98US-0097022
PR	19-AUG-1998	98US-0097218
PR	20-AUG-1998	98US-0097218
PR	24-AUG-1998	98US-0097661
PR	26-AUG-1998	98US-0097951
PR	26-AUG-1998	98US-0097952
PR	26-AUG-1998	98US-0097954
PR	26-AUG-1998	98US-0097955
PR	26-AUG-1998	98US-0097971
PR	26-AUG-1998	98US-0097974
PR	26-AUG-1998	98US-0097978
PR	26-AUG-1998	98US-0097979
PR	26-AUG-1998	98US-0097986
PR	28-AUG-1998	98US-0098014
PR	31-AUG-1998	98US-0098525
PR	16-SEP-1998	98US-0100634
PR	12-JAN-1999	99US-0115565

Baker K, Chén J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI
PI
PI
PI
Wood WI, Yuan J;

Membrane-bound proteins and related nucleotide sequences

The invention provides membrane-bound PRO polypeptides and polynucleotides encoding them. The PRO sequences of the invention were identified based on extracellular domain homology screening. The PRO sequences have homology with proteins including LDL receptors, TIE ligands and various enzymes. The membrane-bound proteins and receptor molecules are useful as pharmaceutical and diagnostic agents. Receptor immunoadhesins, for instance, can be used as therapeutic agents to block receptor-ligand interactions. The membrane-bound proteins can also be employed for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. The PRO encoding sequences are used as hybridization probes, in chromosome and gene mapping and in the generation of antisense RNA and DNA. PRO nucleic acid sequences will also be useful for the preparation of PRO polypeptides, especially by recombinant techniques.

Sequence 919 AA;

```

alignment_scores:
  A      Quality: 286.00      Length: 72
  G      Ratio: 4.333      Gaps: 0
  P      Percent similarity: 91.667      Percent Identity: 73.611

alignment_block:
  AUS-09-049-696-5 x AAY66749 ..
  Alignment 1/1 to: AAY66749 from: 1 to: 919

13 ATAGTTCAATTCGTACAGACAACAAACCAACAAGAGCTCAACAAC 52
14 .....
15 .....
16 .....
17 .....
18 .....
19 .....
20 .....
21 .....
22 .....
23 .....
24 .....
25 .....
26 .....
27 .....
28 .....
29 .....
30 .....
31 .....
32 .....
33 .....
34 .....
35 .....
36 .....
37 .....
38 .....
39 .....
40 .....
41 .....
42 .....
43 .....
44 .....
45 .....
46 ValValGluPheCysAsnGluLysThrHisAsnGlnGluAlaProSerLe 262
47 .....
48 .....
49 .....
50 .....
51 .....
52 .....
53 _GCAAAATCAAAAATGCCAATCCGCAAGCATGGGAAGTGATCCCGTGATT 102
54 .....
55 .....
56 .....
57 .....
58 .....
59 .....
60 .....
61 .....
62 uGlnAsnIleLysCysAsnPheArgSerThrTrpGluValIleSerAsnS 279
63 .....
64 .....
65 .....
66 .....
67 .....
68 .....
69 .....
70 .....
71 .....
72 .....
73 .....
74 .....
75 .....
76 .....
77 .....
78 .....
79 .....
80 .....
81 .....
82 .....
83 .....
84 .....
85 .....
86 .....
87 .....
88 .....
89 .....
90 .....
91 .....
92 .....
93 .....
94 .....
95 .....
96 .....
97 .....
98 .....
99 .....
100 .....
101 .....
102 .....
103 CTGAGGACTTTAGAAAAACCACTCCATGATGACACACAGCCACCAATGCC 152
104 .....
105 .....
106 .....
107 .....
108 .....
109 .....
110 .....
111 .....
112 .....
113 .....
114 .....
115 .....
116 .....
117 .....
118 .....
119 .....
120 .....
121 .....
122 .....
123 .....
124 .....
125 .....
126 .....
127 .....
128 .....
129 .....
130 .....
131 .....
132 .....
133 .....
134 .....
135 .....
136 .....
137 .....
138 .....
139 .....
140 .....
141 .....
142 .....
143 .....
144 .....
145 .....
146 .....
147 .....
148 .....
149 .....
150 .....
151 .....
152 .....
153 ACCTTCATTCGTCGAGATTGGACAAAGAAATGTCGTGTTTACTGCTTGA 202
154 .....
155 .....
156 .....
157 .....
158 .....
159 .....
160 .....
161 .....
162 .....
163 .....
164 .....
165 .....
166 .....
167 .....
168 .....
169 .....
170 .....
171 .....
172 .....
173 .....
174 .....
175 .....
176 .....
177 .....
178 .....
179 .....
180 .....
181 .....
182 .....
183 .....
184 .....
185 .....
186 .....
187 .....
188 .....
189 .....
190 .....
191 .....
192 .....
193 .....
194 .....
195 .....
196 .....
197 .....
198 .....
199 .....
200 .....
201 .....
202 .....
203 .....
204 .....
205 .....
206 .....
207 .....
208 .....
209 .....
210 .....
211 .....
212 .....
213 .....
214 .....
215 .....
216 .....
217 .....
218 .....
219 .....
220 .....
221 .....
222 .....
223 .....
224 .....
225 .....
226 .....
227 .....
228 .....
229 .....
230 .....
231 .....
232 .....
233 .....
234 .....
235 .....
236 .....
237 .....
238 .....
239 .....
240 .....
241 .....
242 .....
243 .....
244 .....
245 .....
246 .....
247 .....
248 .....
249 .....
250 .....
251 .....
252 .....
253 .....
254 .....
255 .....
256 .....
257 .....
258 .....
259 .....
260 .....
261 .....
262 .....
263 .....
264 .....
265 .....
266 .....
267 .....
268 .....
269 .....
270 .....
271 .....
272 .....
273 .....
274 .....
275 .....
276 .....
277 .....
278 .....
279 .....
280 .....
281 .....
282 .....
283 .....
284 .....
285 .....
286 .....
287 .....
288 .....
289 .....
290 .....
291 .....
292 .....
293 .....
294 .....
295 .....
296 .....
297 .....
298 .....
299 .....
300 .....

```

296 ValPheSerLeuLeuLysIleSerGlnArgIleValCysLeuValLeuAs 312
203 CAATCTGGAAGCATG 218
312 pLysSerGlySerMet 317

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB87560

seq_documentation_block:
ID AAB87560 standard; Protein; 919 AA.

AC AAB87560;

DT 15-MAY-2001 (first entry)

DE Human PRO1124.

KW Human; PRO protein; mapping.

XX Homo sapiens.

XX WO200116318-A2.

XX 08-MAR-2001.

XX 24-AUG-2000; 2000WO-US23328.

XX 01-SEP-1999; 99WO-US20111.

XX 15-SEP-1999; 99WO-US21090.

XX 07-DEC-1999; 99US-0169495.

XX 09-DEC-1999; 99US-0170262.

XX 11-JAN-2000; 2000US-0175481.

XX 18-FEB-2000; 2000WO-US04341.

XX 22-FEB-2000; 2000WO-US04342.

XX 01-MAR-2000; 2000WO-US05601.

XX 03-MAR-2000; 2000US-0187202.

XX 25-APR-2000; 2000US-0199397.

XX 22-MAY-2000; 2000WO-US14042.

XX 05-JUN-2000; 2000US-0209832.

XX (GETH) GENENTECH INC.

XX Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;

XX Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;

XX WPI; 2001-183260/18.

XX N-PSDB; AAF92092.

XX Eighty four nucleic acids encoding PRO polypeptides, useful in
XX molecular biology, including use as hybridization probes, and in
XX chromosome and gene mapping.

XX Claim 12; Fig 70; 278pp; English.

XX The present sequence is a human PRO polypeptide (secreted and
XX transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
XX anti-PRO antibodies are useful for preparation of a medicament useful in
XX the treatment of a condition which is responsive to the PRO protein,
XX agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
XX employed as molecular weight markers for protein electrophoresis. The PRO
XX coding sequence has applications in molecular biology, including use as
XX hybridisation probes, and in chromosome and gene mapping.

alignment_scores:
Quality: 286.00 Length: 72
Ratio: 4.333 Gaps: 0

Percent Similarity: 91.667 Percent Identity: 73.611

alignment_block:
US-09-049-696-5 x AAB87560 ..
Align seg 1/1 to: AAB87560 from: 1 to: 919
3 ATAGTTGAATTCGTACAGAAACCAACCAAGAGCTCCAAACAA 52
246 ValValGluPheCysAsnGlnLysThrHisAsnGlnGluAlaProSerIle 262
53 GCAAAATCAAAATGCAATCTCCGAAGCACATGGAGAGTGCCTGATT 102
262 uGlnAsnIleLysCysAsnPheArgSerThrTrpGluValIleSerAsnS 279
103 CTGAGGACTTTAAGAAACCACTCTCTATGACACACACGCCACCAATCCC 152
279 erGluAspPheLysAsnThrIleProMetValThrProProProPro 295
153 ACCTTCTCATTGCTGCAGATTGGACAAAGAAATGTGTGTTAGTCCTTGA 202
296 ValPheSerLeuLysIleSerGlnArgIleValCysLeuValLeuAs 312
203 CAATCTGGAAGCATG 218
312 pLysSerGlySerMet 317

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB65272

seq_documentation_block:
ID AAB65272 standard; Protein; 919 AA.

AC AAB65272;

DT 02-APR-2001 (first entry)

XX Human PRO1124 (UNQ562) protein sequence SEQ ID NO:379.

XX Human; secreted and transmembrane protein; PRO; cytostatic;
XX cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
XX diagnostic assay.

XX Homo sapiens.

XX WO2000073454-A1.

XX 07-DEC-2000.

XX 30-MAR-2000; 2000WO-US08439.

XX 02-JUN-1999; 99WO-US12252.

XX 23-JUN-1999; 99US-0141037.

XX 07-JUL-1999; 99US-0143048.

XX 20-JUL-1999; 99US-0144758.

XX 26-JUL-1999; 99US-0145698.

XX 28-JUL-1999; 99US-0146222.

XX 17-AUG-1999; 99US-0149396.

XX 15-SEP-1999; 99WO-US21090.

XX 15-SEP-1999; 99WO-US21547.

XX 08-OCT-1999; 99US-0158663.

XX 30-NOV-1999; 99WO-US28313.

XX 01-DEC-1999; 99WO-US30095.

XX 16-DEC-1999; 99WO-US30911.

XX 20-DEC-1999; 2000WO-US00219.

XX 05-JAN-2000; 2000WO-US00376.

XX 06-JAN-2000; 2000WO-US03565.

XX 11-FEB-2000; 2000WO-US04341.

XX 18-FEB-2000; 2000WO-US04414.

XX 22-FEB-2000; 2000WO-US04914.

XX 24-FEB-2000; 2000WO-US05004.

XX 02-MAR-2000; 2000WO-US05841.

XX 15-MAR-2000; 2000WO-US06884.

XX 20-MAR-2000; 2000WO-US07377.

(GETH) GENENTECH INC.

Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL; Ferrera N, Fong S, Gerber H, Gerritsen ME, Goddard A, Gadowski PJ; Grimaldi CJ, Gurney AL, Klijavin IJ, Napier MA, Pan J, Paooni NF; Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI; Zhang Z;

WPI: 2001-032160/04.
N-PSDB; AAF44241.

PRO polynucleotides used to produce polypeptides used to target bioactive molecules such as toxins, radiolabels or antibodies, to specific cells, to cause targeted cell death -

Claim 12; Fig 274; 935pp; English.

The present invention describes human secreted and transmembrane proteins. The PRO proteins have cytostatic activity. The PRO proteins can be used for targeted delivery of bioactive molecules, such as toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide sequences, and their fragments, can be used as hybridisation probes, in chromosomal and gene mapping, and in the generation of anti-sense RNA and DNA. They may also be used to produce transgenic animals which are used to develop and screen therapeutically useful reagents. The PRO nucleotide and protein sequence can be used for tissue typing and in treating cancer. Anti-PRO antibodies can be used in diagnostic assays. AAF44270 to AAF44470 represent PCR primers and hybridisation probes used in the isolation of human PRO sequences. AAF44087 to AAF44269 and AAB65154 to AAB65300 represent human PRO polynucleotide and protein sequences given in the exemplification of the present invention.

Sequence 919 AA:

alignment_scores:
Quality: 286.00 Length: 72
Ratio: 4.333 Gaps: 0
Percent Similarity: 91.667 Percent Identity: 73.611

alignment_block:
US-09-049-696-5 x AAB65272 ..

Align seg 1/1 to: AAB65272 from: 1 to: 919

3 ATAGTTGAATTCGTACAGAACAAACCACCAACAAGAAGCTCCAAACAA 52
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
246 ValValGluPheCysAsnGluThrHisAsnGlnGluAlaProSerLe 262

53 GCAAATCAAAAATGCATCTCGAAGCACATGGGAAGTGATCCGTGATT 102
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
262 uGlnAsnIleLysCysAsnPheArgSerThrTrpGluValIleSerAsnS 279

103 CTGAGGACTTAAAGAAACCACTCCTATGACACAGACGCCACCAATCCC 152
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
279 erGluAspPheLysAsnThrIleProMetValThrProProProPro 295

153 ACCTTCTCATGTCGAGATTGACAAAGAATGTGTGTGTAGTCTCTGCA 202
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
296 ValPheSerLeuLeuLysIleSerGlnArgIleValCysLeuValLeuas 312

203 CAATCTGGAAGCATG 218
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
312 pLysSerGlySerMet 317

seq_name: /STDS2/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB45904

seq_documentation_block:
ID AAB45904 standard; Protein; 742 AA.
XX AC AAB45904;
XX DT 21-MAR-2001 (first entry)

Human tumor-associated antigen C42 protein.
Tumor-associated antigen; C42; human; immunogenic; cancer therapy; cytotactic; immunotherapy; vaccine; lung; breast; esophagus.
Homo sapiens;
DE19924199-A1.
30-NOV-2000.
27-MAY-1999; 99DE-1024199.
27-MAY-1999; 99DE-1024199.
(BOEH) BOEHRINGER INGELHEIM INT GMBH.
Adolf G, Heider K, Koenig U, Sommergruber W, Adolf GR, Heider K; Koenig U, Sommergruber W;
WPI: 2001-062549/08.
N-PSDB; AAC82881.

Tumor-associated antigen C42, and related nucleic acids and antibodies, useful in immunotherapy of cancer and for diagnosis -

Claim 1; Page 18; 40pp; German.

This invention describes a novel tumor-associated antigen C42 which has cytostatic activity. The invention also describes (a) immunogenic protein fragments or peptides (I) derived from C42; (b) a pharmaceutical composition for parenteral, topical, oral or local administration containing at least 1 C42 or (I); (c) an isolated DNA (II) encoding C42 or its fragments; (d) a recombinant DNA (III) that contains (II); and (e) antibodies (Ab) raised against C42 or (I). C42 and its fragments induce a humoral immune response and, when presented by major histocompatibility complex molecules, a cellular immune response. C42 and its fragments, are used for immunotherapy of cancers that express C42 (particularly of lung, breast and esophagus), to raise specific antibodies (Ab) and for diagnosis, e.g. to detect induction of an immune response and for optimization of treatment. Ab are used therapeutically (e.g. when conjugated to a cytotoxin) or for diagnosis or monitoring of cancers that express C42. Nucleic acid (II) that encodes C42 can also be used for immunotherapy and cells that express C42 as antitumor vaccines.

Sequence 742 AA;

alignment_scores:
Quality: 232.00 Length: 74
Ratio: 3.932 Gaps: 1
Percent Similarity: 79.730 Percent Identity: 59.459

alignment_block:
US-09-049-696-5 x AAB45904 ..

Align seg 1/1 to: AAB45904 from: 1 to: 742

3 ATAGTTGAATTCGTACAGAACAAACCACCAACAAGAAGCTCCAAACAA 52
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
43 ValValGluPheCysAsnAlaSerThrHisAsnGlnGluAlaProAsnLe 64

53 GCAAATCAAAAATGCATCTCGAAGCACATGGGAAGTGATCCGTGATT 102
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
64 uGlnAsnGlnMetCysSerLeuArgSerAlaTrpAspValIleThrAsps 81

103 CTGAGGACTTAAAGAAACCACTCCTATGACAACACAGCCACCA 146
|| ||||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
81 erAlaaspPheHisSerPhePrometAsnGlyThrGluLeuProPro 97

147 AATCCACCTTCTCATGTCGAGATTGGACAAAGAATGTGTGTGTAGT 196
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Percent Similarity: 79.730 Percent Identity: 58.108

alignment_block:

US-09-049-696-5 x AAB11324

Align seg 1/1 to: AAB11324 from: 1 to: 592

```
3 ATAGTTGAATTCGTACAGACAAACACACAAAGAGCTCCAAACAA 52
  ::::::::::::::::::::|
249 ValValGluPheCysAsnAlaSerThrHisAsnGlnGluAlaProAsnLe 265
  ::::::::::::::::::::|
53 GCAAAATCAAAAATGCAATCTCCGAACACATGGGAAGTGCATCGTGATT 102
  ::::::::::::::::::::|
265 uGlnAsnGlnMetCysSerLeuArgSerAlaTrpAspValIleThrAsps 282
  ::::::::::::::::::::|
103 CTGAGGACTTTAAGAAACCACTCCTATG.....ACAACACAGCCACCA 146
  ::::::::::::::::::::|
282 erAlaAspPheHisHisSerPheProMetAsnGlyThrGluLeuProPro 298
  ::::::::::::::::::::|
147 AATCCACCTTCTCATTTGCTGCAGATTGGACAAAGAATTGTGTTTGTAGT 196
  ::::::::::::::::::::|
299 ProProThrPheSerLeuValGluAlaGlyAspLysValValCysLeuVa 315
  ::::::::::::::::::::|
197 CCTTGACAAATCTGGAAGCATG 218
  ::::::::::::::::::::|
315 lLeuAspValSerSerLysMet 322
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AA41040

seq_documentation_block:

ID AAY41040 standard; protein; 791 AA.

XX AC AAY41040;

XX DT 07-DEC-1999 (first entry)

XX DE Human lung tumor antigen L762P variant 2.

XX KW Human; lung tumor; lung cancer; T cell stimulation.

XX OS Homo sapiens.

XX PN W09947674-A2.

XX PD 23-SEP-1999.

XX PF 17-MAR-1999; 99WO-US05798.

XX PR 18-MAR-1998; 98US-0040802.

XX PR 18-MAR-1998; 98US-0040984.

XX PR 27-JUL-1998; 98US-0123912.

XX PR 27-JUL-1998; 98US-0123933.

XX PA (CORI-) CORIXA CORP.

XX PI Reed SG, Wang T;

XX DR WPI; 1999-571839/48.

XX DR N-PSDB; AA224658.

XX PT New isolated lung tumor polynucleotides, used to develop products for

XX PT the treatment, prevention and monitoring the progression of lung cancer

XX PS Example 3; Page 145-147; 148pp; English.

XX CC The invention provides isolated human lung tumor nucleic acids and
XX CC polypeptides. The polypeptides can be used for the treatment of lung
XX CC cancer. The polypeptides and polynucleotides can be used to stimulate T
XX CC cells or antigen presenting cells for use in the treatment of lung
XX CC cancer. The polypeptides and monoclonal antibodies specific for the
XX CC polypeptides can also be used to inhibit the development of lung cancer.
XX CC Agents which bind the polypeptides can be used for detecting lung cancer

CC and for monitoring the progression of lung cancer.

XX SQ

Sequence 791 AA;

alignment_scores:

Quality: 229.00 Length: 74
Ratio: 3.881 Gaps: 1
Percent Similarity: 79.730 Percent Identity: 58.108

alignment_block:

US-09-049-696-5 x AAY41040

Align seg 1/1 to: AAY41040 from: 1 to: 791

```
3 ATAGTTGAATTCGTACAGACAAACACACAAAGAGCTCCAAACAA 52
  ::::::::::::::::::::|
249 ValValGluPheCysAsnAlaSerThrHisAsnGlnGluAlaProAsnLe 265
  ::::::::::::::::::::|
53 GCAAAATCAAAAATGCAATCTCCGAACACATGGGAAGTGCATCGTGATT 102
  ::::::::::::::::::::|
265 uGlnAsnGlnMetCysSerLeuArgSerAlaTrpAspValIleThrAsps 282
  ::::::::::::::::::::|
103 CTGAGGACTTTAAGAAACCACTCCTATG.....ACAACACAGCCACCA 146
  ::::::::::::::::::::|
282 erAlaAspPheHisHisSerPheProMetAsnGlyThrGluLeuProPro 298
  ::::::::::::::::::::|
147 AATCCACCTTCTCATTTGCTGCAGATTGGACAAAGAATTGTGTTTGTAGT 196
  ::::::::::::::::::::|
299 ProProThrPheSerLeuValGluAlaGlyAspLysValValCysLeuVa 315
  ::::::::::::::::::::|
197 CCTTGACAAATCTGGAAGCATG 218
  ::::::::::::::::::::|
315 lLeuAspValSerSerLysMet 322
```

This Page Blank (uspto)

OM of: US-09-049-696-4 to: SPTREMBL_17:* out_format : pfs
Date: Mar 30, 2002 2:46 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+nt2p.model -DEV=xlp
-Q=/cgn2.1/USPTO_spool/US09049696/runat_28032002.145238.2053/app_query.fasta_1.12579
-DB=SPTREMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcp
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696_ECGN1_1_705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-4
Query length: 181
Database: SPTREMBL_17*
Database sequences: 473505
Database length: 14627329
Search time (sec): 805.760000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
sp_human:Q95151	+ 324.00	752.76	1.3e-34	914	! Documentation
sp_human:Q9UPC6	+ 324.00	752.76	1.3e-34	914	! Q9UPC6 homo sapiens (human). ca
sp_human:Q9UNF6	+ 324.00	752.76	1.3e-34	914	! Q9UNF6 homo sapiens (human). ca
sp_mammal:Q9RUB5	+ 224.00	513.37	2.0e-21	917	! Q9RUB5 sus scrofa (pig). epithe
sp_rodent:Q98826	+ 210.00	479.89	2.0e-19	913	! Q98826 mus musculus (mouse). gc
sp_rodent:Q9B726	+ 210.00	479.89	2.0e-19	913	! Q9B726 mus musculus (mouse). ch
sp_human:Q9UNF7	+ 180.00	408.05	2.0e-15	917	! Q9UNF7 homo sapiens (human). ca
sp_rodent:Q98860	+ 170.00	384.24	4.4e-14	902	! Q98860 mus musculus (mouse). ch
sp_rodent:Q9OX15	+ 170.00	384.23	4.4e-14	902	! Q9OX15 mus musculus (mouse). ca
sp_rodent:Q9R070	+ 169.00	381.84	6.0e-14	902	! Q9R070 mus musculus (mouse). ca
sp_rodent:Q9R0R4	+ 169.00	381.84	6.0e-14	902	! Q9R0R4 mus musculus (mouse). en
sp_mammal:O18744	+ 160.00	367.49	9.5e-13	342	! O18744 bos taurus (bovine). lu
sp_mammal:O18742	+ 160.00	361.24	9.5e-13	794	! O18742 bos taurus (bovine). lu
sp_mammal:O18743	+ 160.00	361.00	9.5e-13	820	! O18743 bos taurus (bovine). lu
sp_mammal:O18741	+ 160.00	360.27	9.5e-13	905	! O18741 bos taurus (bovine). lu
sp_human:Q9UQC9	+ 156.50	351.59	2.8e-12	943	! Q9UQC9 homo sapiens (human). ch
sp_human:Q9Y6N2	+ 156.50	351.59	2.8e-12	943	! Q9Y6N2 homo sapiens (human). ca
sp_human:Q9Y6N3	+ 156.00	359.90	3.4e-12	262	! Q9Y6N3 homo sapiens (human). ch
sp_virus:Q40959	+ 77.00	154.67	0.1055	2301	! Q40959 little cherry closterov
sp_rodent:Q9O2Y0	+ 71.00	146.30	0.6919	1026	! Q9O2Y0 mus musculus (mouse). F
sp_organelle:Q9GE29	+ 68.00	147.77	1.84	320	! Q9GE29 collomia linearis (narrow
sp_invertebrate:Q26008	+ 66.00	126.60	3.05	2910	! Q26008 plasmodium falciparum
sp_virus:Q9WLG1	- 65.00	154.36	5.05	50	! Q9WLG1 human coxsackievirus b6.
sp_organelle:Q9GF03	- 65.00	140.54	4.62	322	! Q9GF03 allopythium divaricatum
sp_virus:Q989280	- 65.00	140.20	4.61	337	! Q989280 human coxsackievirus b6.
sp_virus:Q9YX63	- 65.00	133.01	4.40	888	! Q9YX63 human coxsackievirus b6.
sp_virus:Q9QL88	- 65.00	126.34	4.21	2184	! Q9QL88 human coxsackievirus b6.
sp_virus:Q9QAH1	- 65.00	126.34	4.21	2184	! Q9QAH1 human coxsackievirus b6.
sp_organelle:Q9GEZ7	- 64.00	138.10	6.27	324	! Q9GEZ7 gallia leptalea. nadh del
sp_plant:Q9SW08	- 63.50	132.62	7.11	577	! Q9SW08 arabidopsis thaliana (md
sp_plant:Q9M9X9	- 63.50	128.64	6.93	987	! Q9M9X9 arabidopsis thaliana (md
sp_organelle:Q9GF04	- 63.00	135.50	8.52	333	! Q9GF04 gallia latifolia. nadh de
sp_organelle:Q9GEY3	- 63.00	135.50	8.52	333	! Q9GEY3 navarretia intertexta. n
sp_virus:Q9PYE3	- 63.00	128.54	8.14	851	! Q9PYE3 human coxsackievirus b3.
sp_invertebrate:Q22601	+ 63.00	128.07	8.11	907	! Q22601 caenorhabditis elegans
sp_virus:Q9E7C1	- 63.00	127.83	8.10	936	! Q9E7C1 human coxsackievirus b3.
sp_virus:Q66331	- 63.00	127.14	8.07	1028	! Q66331 human coxsackievirus b3.
sp_virus:Q66338	- 63.00	121.54	7.78	2185	! Q66338 human coxsackievirus b3.
sp_virus:Q9E7C4	- 63.00	121.54	7.78	2185	! Q9E7C4 human coxsackievirus b3.

sp_virus:Q9E7C3	-	63.00	121.54	7.78	2185	! Q9E7C3 human coxsackievirus
sp_virus:Q9E7C2	-	63.00	121.54	7.78	2185	! Q9E7C2 human coxsackievirus
sp_invertebrate:Q9XTE6	-	62.00	132.39	11.52	367	! Q9XTE6 caenorhabditis eleg
sp_invertebrate:Q9VAM2	-	61.50	125.28	12.93	814	! Q9VAM2 drosophila melanoga
sp_virus:Q9IX00	-	61.00	146.44	17.44	40	! Q9IX00 human coxsackievirus b
seq_name: sp_human:Q95151						
seq_documentation_block:						
ID	Q95151	PRELIMINARY;	PRT;	914	AA.	
AC	Q95151;					
DC	01-MAY-1999 (TREMBLrel. 10, Created)					
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)					
DE	01-JUN-2001 (TREMBLrel. 17, Last annotation update)					
DE	CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.					
GN	HCJCAL.					
GS	Homo sapiens (Human).					
CC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
OC	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.					
NCBI_TaxID	9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=SMALL INTESTINE;					
RX	MEDLINE=99047526; PubMed=9828122;					
RA	Gruber A.D., Elble R.C., Ji H.L., Schreuer K.D., Fuller C.M.,					
RA	Pauli B.U.;					
RT	"Genomic cloning, molecular characterization, and functional analysis					
RT	of the human CLCA1, the first human member of the family of Ca2+-activated					
RT	Cl-channel proteins."					
RL	Genomics 54:200-214(1998).					
DR	EMBL; AF039400; AAC95428.1;					
DR	InterPro; IPR000131; ATPase_gamma.					
DR	PROSITE; P500153; ATPASE_GAMMA; UNKNOWN_1.					
DR	PROSITE; P500234; VWFA; 1.					
DR	SMART; SM00327; VWA; 1.					
SQ	SEQUENCE 914 AA; 100174 MW; 7066440C46557FA3 CRC64;					
alignment_scores:						
Quality:	324.00	Length:	60			
Ratio:	5.400	Gaps:	0			
Percent Similarity:	100.000	Percent Identity:	100.000			
alignment_block:						
US-09-049-696-4 x Q95151						
Align seg 1/1 to: Q95151 from: 1 to: 914						
1	AAAGATGCACATTCATTAAGTAAGTACAGGACTCTATGAAAAGGATGTGA	51				
208	LysargCysThrPheAsnLysValThrGlyLeuTrGluLysGlyCysG	224				
52	TTTTGTCTCAATCCCGCCAGCAGGAGAGGCTTCTATATATTTTCAC	101				
224	upheValLeuGlnSerArgGlnThrGluLysAlaSerIleMetPheAlaG	241				
102	AACATGTTGATTCATATAGTTGAATTTCTACAGAACAAACACACAAA	151				
241	lnHisValAspSerIleValGluPheCysThrGluGlnAsnHisAsnLys	257				
152	GAAGCTCAACACAGCAAAATCAAAATGC	181				
258	GluAlaProAsnLysGlnAsnGlnLysCys	267				
seq_name: sp_human:Q9UPC6						
seq_documentation_block:						
ID	Q9UPC6	PRELIMINARY;	PRT;	914	AA.	
AC	Q9UPC6;					
DT	01-MAY-2000 (TREMBLrel. 13, Created)					
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)					
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)					

Tue Apr 2 09:40:03 2002

us-09-049-696-4.rspt

```

DE HCLCAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9047526; PubMed=9828122;
RA Gruber A.D., Elble R.C., Ji H.L., Schreier K.D., Fuller C.M.,
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
RT of human CICAL, the first human member of the family of Ca2+-activated
RT Cl-channel proteins."
RL Genomics 54:200-214(1998).
DR EMBL: AF039401; AAC95429.1; .
DR InterPro: IPR000131; ATPase_gamma.
DR DR PROSITE: PS00153; ATPase_GAMMA; UNKNOWN_1.
DR DR PROSITE: PS50234; VWFA; 1.
DR SMART: SM00327; VWFA; 1.
SQ SEQUENCE 914 AA; 100160 MW; 7761362B66752AA3 CRC64;

alignment_scores:
Quality: 324.00 Length: 60
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-4 x Q9UNF6 ..
Align seg 1/1 to: Q9UNF6 from: 1 to: 914

2 AAAGATGCACATTCATTAAGTAACAGGACTCTATGAAAAAGGATGTGA 51
|||||
208 LysArgCysThrPheAsnLysValThrGlyLeuYrGluLysGlyCysG1 224
|||||
52 GTTCTCTCCCAATCCCGCAGAGGAGGCTTCTATATGTTGCAC 101
|||||
224 uPheValLeuGlnSerArgGlnThrGluLysAlaSerIleMetPheAlaG 241
|||||
102 AACATGTTGATTTCTATAGTTGAATTCGTACAGACAAACCAACAA 151
|||||
241 LhHisValAspSerIleValGluPheCysThrGluGlnAsnHisAsnLys 257
|||||
152 GAAGCTCCAAACAAAGCAAAATCAAAATGC 181
|||||
258 GluAlaProAsnLysGlnAsnGlnLysCys 267

seq_name: sp_mammal:Q9TUB5
seq_documentation_block:
ID Q9TUB5 PRELIMINARY; PRT; 917 AA.
AC Q9TUB5
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN.
GN AECC.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9623;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=ILEAL MUCOSA;
RA Gaspar K.J., Gabriel S.E., Racette K.J., Forsyth G.W.;
RT "A CDNA involved in porcine exocrine chloride conductance."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF095584; AAF00077.1; .
DR InterPro: IPR000131; ATPase_gamma.
DR DR PROSITE: PS00153; ATPase_GAMMA; UNKNOWN_1.
DR DR PROSITE: PS50234; VWFA; 1.
DR SMART: SM00327; VWFA; 1.
SQ SEQUENCE 917 AA; 100735 MW; 649F41944F7EAD19 CRC64;

alignment_scores:
Quality: 224.00 Length: 60
Ratio: 4.392 Gaps: 0
Percent Similarity: 85.000 Percent Identity: 65.000

alignment_block:
US-09-049-696-4 x Q9TUB5 ..
Align seg 1/1 to: Q9TUB5 from: 1 to: 917

```

```
2 AAAAGATGCACATTCATAAAGTAACAGGACTCTATGAAAGGATGTGA 51
||| ||| :|||:|||||:|||||:|||||:|||||
208 LysProCysArgAlaAspArgValThrGlyLeuPheGlnLysGluCysG1 224
52 GTTGTCTCAATCCCGCCAGGAGGCTTCTATATGTTTCAC 101
|||||: ||| :|||:|||||:|||||:|||||:|||||
224 uPheIleProAspGlnSerGluLysAlaSerIleMetPheAlaG 241
102 AACATGTTGATCTCTATAGTTGAATTCGTACAGAACAAACCAACAAA 151
|| :|||:|||||:|||||:|||||:|||||:|||||
241 InSerIleAspThrValGluPheCysLysGluLysAsnHisAsnLys 257
152 GAAGCTCAAAACAGCAAAATCAAAATGTC 181
|||||:|||||:|||||:|||||:|||||
258 GluAlaProAsnAspGlnAsnGlnLysCys 267
```

seq_name: sp_rodent:088826

```
seq_documentation_block:
ID O88826 PRELIMINARY; PRT; 913 AA.
AC O88826;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GOB-5 PROTEIN.
GN CLCA3 OR GOB-5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RA Komiyu T., Tanigawa Y., Hirohashi S.;
RT "Cloning of a gene, gob-5, which is expressed in intestinal goblet
cells in mice."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017156; BAA33743.1; -.
DR MGD; MGI:1346342; Clca3.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 913 AA; 100070 MW; A7FA2F9E1089806D CRC64;
```

```
alignment_scores:
Quality: 210.00 Length: 59
Ratio: 3.962 Gaps: 0
Percent Similarity: 89.831 Percent Identity: 61.017

alignment_block:
US-09-049-696-4 x O88826 ..
Align seg 1/1 to: O88826 from: 1 to: 913
```

```
5 AGATGCACATTCATAAAGTAACAGGACTCTATGAAAGGATGTGAGTT 54
:|||||:|||||:|||||:|||||:|||||:|||||
210 LysCysValIleAspArgValThrGlyLeuTyrlLysAspAsnCysValPh 226
55 TGTCTCAATCCCGCCAGGAGGCTTCTATATGTTTCACAAAC 104
||| :|||:|||||:|||||:|||||:|||||
226 eValProAspProHisGlnAsnGluLysAlaSerIleMetPheAsnGlnA 243
105 ATGTTGATTCCTATAGTTGAATTCGTACAGAACAAACCAACAAAGAA 154
:|||||:|||||:|||||:|||||:|||||:|||||
243 snIleAsnSerValGluPheCysThrGluLysAsnHisAsnGlnGlu 259
155 GCTCCAAACAGCAAAATCAAAATGTC 181
|||||:|||||:|||||:|||||:|||||
260 AlaProAsnAspGlnAsnGlnArgCys 268
```

seq_name: sp_rodent:Q9D726

```
seq_documentation_block:
ID Q9D726 PRELIMINARY; PRT; 913 AA.
AC Q9D726;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CHLORIDE CHANNEL CALCIUM ACTIVATED 3.
GN CLCA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=STOMACH;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Flöschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK008659; BAB25815.1; -.
DR MGD; MGI:1346342; Clca3.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 913 AA; 100111 MW; 491E8584B06D9A89 CRC64;
```

```
alignment_scores:
Quality: 210.00 Length: 59
Ratio: 3.962 Gaps: 0
Percent Similarity: 89.831 Percent Identity: 61.017

alignment_block:
US-09-049-696-4 x Q9D726 ..
Align seg 1/1 to: Q9D726 from: 1 to: 913
5 AGATGCACATTCATAAAGTAACAGGACTCTATGAAAGGATGTGAGTT 54
:|||||:|||||:|||||:|||||:|||||:|||||
210 LysCysValIleAspArgValThrGlyLeuTyrlLysAspAsnCysValPh 226
55 TGTCTCAATCCCGCCAGGAGGCTTCTATATGTTTCACAAAC 104
||| :|||:|||||:|||||:|||||:|||||
226 eValProAspProHisGlnAsnGluLysAlaSerIleMetPheAsnGlnA 243
105 ATGTTGATTCCTATAGTTGAATTCGTACAGAACAAACCAACAAAGAA 154
:|||||:|||||:|||||:|||||:|||||:|||||
248 snIleAsnSerValGluPheCysThrGluLysAsnHisAsnGlnGlu 259
155 GCTCCAAACAGCAAAATCAAAATGTC 181
|||||:|||||:|||||:|||||:|||||
```


[illegible]

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elbie R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001264; AAB86532.1; -
SQ SEQUENCE 342 AA; 39037 MW; 6A491480B4BED01A CRC64;

alignment_scores:
  Quality: 160.00      Length: 58
  Ratio: 3.902        Gaps: 0
Percent Similarity: 70.690 Percent Identity: 51.724

alignment_block:
US-09-049-696-4 x O18744
Align seg 1/1 to: O18744 from: 1 to: 342
8 TGCACATTCAATAAGTAACAGGACTCTATGAAAGAGGTGAGTTTGT 57
   ||| : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
211 CysArgArgAspSerGlnThrGlyLeuTyrGluAlaLysCysThrPheLe 227
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
58 TCTCCAATCCGCCAGCAGGAGAGGCTTCTATATGTTTGCACACATG 107
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
227 uProLysLysSerGlnThrAlaLysGluSerIleMetPheMetProSerL 244
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
108 TTGATTTCTATAGTTGAATTCGTACAGAACAAACACCAACAAAGAGCT 157
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
244 euHisSerValThrGluPheCysThrGluLysThrHisAsnThrGluAla 260
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
158 CCAACAAGCAAAATCAAAATGC 181
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
261 ProAsnLeuGlnAsnLysMetCys 268

seq_name: sp_mammal:O18744
seq_documentation_block:
ID O18744 PRELIMINARY; PRT; 820 AA.
AC O18743;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
CC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elbie R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001263; AAB86531.1; -
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR SEQUENCE 820 AA; 91464 MW; 4E01330AEF5A1494 CRC64;

alignment_scores:
  Quality: 160.00      Length: 58
  Ratio: 3.902        Gaps: 0
Percent Similarity: 70.690 Percent Identity: 51.724

alignment_block:
US-09-049-696-4 x O18743
Align seg 1/1 to: O18743 from: 1 to: 820
8 TGCACATTCAATAAGTAACAGGACTCTATGAAAGAGGTGAGTTTGT 57
   ||| : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
211 CysArgArgAspSerGlnThrGlyLeuTyrGluAlaLysCysThrPheLe 227
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
58 TCTCCAATCCGCCAGCAGGAGAGGCTTCTATATGTTTGCACACATG 107
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
227 uProLysLysSerGlnThrAlaLysGluSerIleMetPheMetProSerL 244
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
108 TTGATTTCTATAGTTGAATTCGTACAGAACAAACACCAACAAAGAGCT 157
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
244 euHisSerValThrGluPheCysThrGluLysThrHisAsnThrGluAla 260
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
158 CCAACAAGCAAAATCAAAATGC 181
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
261 ProAsnLeuGlnAsnLysMetCys 268

seq_name: sp_mammal:O18743
seq_documentation_block:
ID O18743 PRELIMINARY; PRT; 820 AA.
AC O18743;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
CC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elbie R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001263; AAB86531.1; -
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR SEQUENCE 820 AA; 91464 MW; 4E01330AEF5A1494 CRC64;
```

```
US-09-049-696-4 x O18742
Align seg 1/1 to: O18742 from: 1 to: 794
8 TGCACATTCAATAAGTAACAGGACTCTATGAAAGAGGTGAGTTTGT 57
   ||| : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
211 CysArgArgAspSerGlnThrGlyLeuTyrGluAlaLysCysThrPheLe 227
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
58 TCTCCAATCCGCCAGCAGGAGAGGCTTCTATATGTTTGCACACATG 107
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
227 uProLysLysSerGlnThrAlaLysGluSerIleMetPheMetProSerL 244
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
108 TTGATTTCTATAGTTGAATTCGTACAGAACAAACACCAACAAAGAGCT 157
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
244 euHisSerValThrGluPheCysThrGluLysThrHisAsnThrGluAla 260
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
158 CCAACAAGCAAAATCAAAATGC 181
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
261 ProAsnLeuGlnAsnLysMetCys 268

seq_name: sp_mammal:O18742
seq_documentation_block:
ID O18742 PRELIMINARY; PRT; 820 AA.
AC O18742;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
CC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elbie R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001263; AAB86531.1; -
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR SEQUENCE 820 AA; 91464 MW; 4E01330AEF5A1494 CRC64;

alignment_scores:
  Quality: 160.00      Length: 58
  Ratio: 3.902        Gaps: 0
Percent Similarity: 70.690 Percent Identity: 51.724

alignment_block:
US-09-049-696-4 x O18743
Align seg 1/1 to: O18743 from: 1 to: 820
8 TGCACATTCAATAAGTAACAGGACTCTATGAAAGAGGTGAGTTTGT 57
   ||| : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
211 CysArgArgAspSerGlnThrGlyLeuTyrGluAlaLysCysThrPheLe 227
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
58 TCTCCAATCCGCCAGCAGGAGAGGCTTCTATATGTTTGCACACATG 107
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
227 uProLysLysSerGlnThrAlaLysGluSerIleMetPheMetProSerL 244
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
108 TTGATTTCTATAGTTGAATTCGTACAGAACAAACACCAACAAAGAGCT 157
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
244 euHisSerValThrGluPheCysThrGluLysThrHisAsnThrGluAla 260
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
158 CCAACAAGCAAAATCAAAATGC 181
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| :
261 ProAsnLeuGlnAsnLysMetCys 268

seq_name: sp_mammal:O18743
seq_documentation_block:
ID O18743 PRELIMINARY; PRT; 820 AA.
AC O18743;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
CC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elbie R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001263; AAB86531.1; -
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR SEQUENCE 820 AA; 91464 MW; 4E01330AEF5A1494 CRC64;
```


This page Blank (uspto)

OM of: US-09-049-696-4 to: SwissProt_39:* out_format : pfs

Date: Mar 30, 2002 2:51 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+nt2p.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool/US09049696/runat_28032002.145238.2085/app_query.fasta.1.12579
-DB=SwissProt_39 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DEPOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcp
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1_165 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-4

Query length: 181

Database: SwissProt_39.*

Database sequences: 100059

Database length: 3664827

Search time (sec): 306.030000

score_list:	Sequence	Strd Orig	ZScore	EScore	Len	Documentation
	SwissProt_39:ECUC_BOVIN +	165.00	359.19	2.7e-13	903	! P54281 bos taurus (bovine).
	SwissProt_39:POLG_CXB3N -	63.00	116.12	3.91	2185	! P03313 c genome polyprotein
	SwissProt_39:POLG_CXB3N -	63.00	116.12	3.91	2185	! P06282 c genome polyprotein
	SwissProt_39:A33_PLEWA +	62.50	125.07	4.35	624	! Q02084 pleurodeles waltii (i
	SwissProt_39:SR68_HUMAN +	62.00	123.98	5.04	619	! Q9ubh9 homo sapiens (human).
	SwissProt_39:SR68_CANFA +	62.00	123.94	5.04	622	! Q00004 canis familiaris (dog
	SwissProt_39:O6B1_HUMAN -	61.00	127.21	6.62	311	! O95007 homo sapiens (human).
	SwissProt_39:INX6_DROME -	61.00	123.70	6.72	481	! Q9vri8 drosophila melanogast
	SwissProt_39:DPOL_HPBBE +	61.00	119.72	6.84	788	! P13846 heron hepatitis b vir
	SwissProt_39:POLG_CXB4E +	61.00	111.50	7.08	2183	! Q86887 c genome polyprotein
	SwissProt_39:POLG_CXB4J -	61.00	111.50	7.08	2183	! P08232 c genome polyprotein
	SwissProt_39:FLIM_BUCAL +	59.50	123.64	10.34	315	! P57182 buchera aphidicola (
	SwissProt_39:A2A2_RAT +	59.00	113.69	12.45	937	! P18484 rattus norvegicus adapt
	SwissProt_39:A2A2_MOUSE +	59.00	113.69	12.45	938	! P17427 mus musculus adaptor
	SwissProt_39:A2A2_HUMAN +	59.00	113.68	12.45	939	! O94973 h adapter-related pro
	SwissProt_39:TYCA_BACBR +	59.00	112.49	12.51	1088	! P09095 bacillus brevis. tyr
	SwissProt_39:CYB1_ESCCA -	58.50	117.82	14.12	487	! O64899 eschscholzia califor
	SwissProt_39:OX40_RAT +	58.00	121.39	16.05	271	! P15725 rattus norvegicus (rat)
	SwissProt_39:DBDR_XENLA -	58.00	117.17	16.34	457	! P42290 xenopus laevis (afri
	SwissProt_39:NUSM_TRYBB -	58.00	115.11	16.48	590	! P04540 trypanosoma brucei br
	SwissProt_39:STB6_YEAST -	58.00	108.80	16.93	1290	! P12866 saccharomyces cerevi
	SwissProt_39:YBET_YEAST -	58.00	108.25	16.97	1381	! P34216 saccharomyces cerevi
	SwissProt_39:ADMR_MOUSE -	57.50	117.19	18.85	395	! P43142 mus musculus (mouse).
	SwissProt_39:Z228_HUMAN -	57.50	110.43	19.40	913	! Q9uju3 homo sapiens (human).
	SwissProt_39:CTA5_RABIT -	57.50	107.81	19.62	1264	! P04144 oryctolagus cuniculu
	SwissProt_39:POLG_CXB1J -	57.00	102.25	23.19	2182	! P08291 c genome polyprotein
	SwissProt_39:POLG_EC12T -	57.00	102.21	23.20	2193	! Q66575 e genome polyprotein
	SwissProt_39:VL02_VACCV -	56.50	127.08	24.08	87	! P20843 vaccinia virus (strain
	SwissProt_39:VL02_VARV -	56.50	127.08	24.08	87	! P33041 variola virus. protein
	SwissProt_39:PD22_RAT -	56.50	117.45	25.09	287	! P47816 rattus norvegicus (rat)
	SwissProt_39:PD22_MOUSE -	56.50	116.02	25.24	343	! P46718 mus musculus (mouse).
	SwissProt_39:Y316_MYCE -	56.50	115.43	25.31	369	! P47558 mycoplasma genitalium
	SwissProt_39:BRCC2_HUMAN +	56.50	97.47	27.32	3418	! P51587 homo sapiens (human).
	SwissProt_39:GPRF_CERAE +	56.00	114.47	29.33	360	! O18982 cercopithecus aethiop
	SwissProt_39:CKF6_HUMAN -	56.00	114.16	29.37	374	! P51684 homo sapiens (human).
	SwissProt_39:BLMH_CHICK -	56.00	112.58	29.57	455	! P87362 gallus gallus (chicke
	SwissProt_39:PMG2_ECOLI +	55.50	117.47	33.42	215	! P36942 escherichia coli. pro
	SwissProt_39:CB82_ESCCA -	55.50	110.86	34.38	488	! O64900 eschscholzia califor
	SwissProt_39:PROD_HUMAN +	55.50	110.41	34.44	516	! O43272 homo sapiens (human).
	SwissProt_39:LOX5_RAT +	55.50	108.28	34.76	672	! P12527 rattus norvegicus (rat)

SwissProt_39:METH_SNY3 + 55.50 103.64 35.45 1195 ! Q55786 synechocystis sp.
SwissProt_39:ABC3_HUMAN - 55.50 100.77 35.88 1704 ! Q99758 homo sapiens (hum
SwissProt_39:Y06N_BPT4 - 55.00 121.02 38.00 120 ! P32274 bacteriophage t4. h
SwissProt_39:UL20_HCMVA - 55.00 112.62 39.38 340 ! P16758 human cytomegalovi
SwissProt_39:POLG_SVDVH - 55.00 97.61 41.98 2185 ! P16604 s genome polyprot
seq_name: SwissProt_39:ECUC_BOVIN
seq_documentation_block:
ID ECUC_BOVIN STANDARD; PRT; 903 AA.
AC P54281;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN (CALCIUM-ACTIVATED CHLORIDE
CHANNEL).
DE Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN 111
RP SEQUENCE FROM N.A.
RC TISSUE=Trachea;
RX MEDLINE=96125078; PubMed=8537359;
RA Cunningham S.A., Awayda M.S., Bublten J.K., Ismailov I.I.,
RA Arrate M.P., Berdiev B.K., Benos D.J., Fuller C.M.;
RT "Cloning of an epithelial chloride channel from bovine trachea.";
RL J.Biol. Chem. 270:31016-31026(1995).
CC ! FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE
CC SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME;
CC MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND
CC TRANSEPITHELIAL TRANSPORT.
CC ! SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC ! TISSUE SPECIFICITY: TRACHEA.
CC ! PTM: PHOSPHORYLATED BY CAM-KINASE II.
CC ! SIMILARITY: CONTAINS 1 WFVA DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: U36445; AAC48511.1; -
InterPro: IPR002035; WFVA.
SMART: SM00327; WVA; 1.
PROSITE: PS0234; WFVA; 1.
Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
Calcium channel; Phosphorylation; Glycoprotein.
TRANSMEM 7 27 POTENTIAL.
TRANSMEM 331 351 POTENTIAL.
TRANSMEM 617 637 POTENTIAL.
TRANSMEM 883 903 POTENTIAL.
DOMAIN 308 476 WFVA.
FT CARBOHYD 75 75 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 535 535 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 688 688 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 811 811 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 816 816 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 842 842 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 857 857 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 903 AA; 100305 MW; 9742AB1590908AEC CRC64;

alignment_scores:

Quality: 165.00 Length: 58

Ratio: 3.929 Gaps: 0
Percent Similarity: 72.414 Percent Identity: 51.724

alignment_block:

US-09-049-696-4 x ECLC_BOVIN ..

Align seg 1/1 to: ECLC_BOVIN from: 1 to: 903

```
8 TCACATTCATTAAGTACAGACTCTATGAAAGAGATGTGAGTTTGT 57
|||||
210 CysArgArgSerGlnThrGlyLeuThrGluAlaLysCysThrPheI 226
|||||
58 TCTCCAAATCCGCCACGACGAGAGCGTCTTATAATGTTGCACACATG 107
|||||
226 eProGluLysSerGlnThrAlaArgGluSerIleMetPheMetGlnSerL 243
|||||
108 TTGATTCTTATAGTTGATTTCTTACAGAACAAACACACACAAAGAAGCT 157
|||||
243 euHisSerValThrGluPheCysThrGluLysThrHisAsnValGluAla 259
|||||
158 CCAAAACAGCAAAATCAAAATGCG 181
|||||
260 ProAsnLeuGlnAsnLysMetCys 267
```

seq_name: SwissProt_39:POLG_CXB3N

seq_documentation_block:

```
ID POLG_CXB3N STANDARD; PRT: 2185 AA.
AC P08313; Q66322; Q66323; Q66324; Q66325; Q66326; Q66327; Q66328;
AC Q83744;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEIN VP4 (PIA); COAT PROTEIN VP2
DE (PIB); COAT PROTEIN VP3 (PIC); COAT PROTEIN P2B; CORE PROTEIN 2A
DE (EC 3.4.22.29) (P2A); CORE PROTEIN P2B; CORE PROTEIN P2C; CORE PROTEIN
DE P3A; GENOME-LINKED PROTEIN VPG (P3B); PICORNAIN 3C (EC 3.4.22.28)
DE (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (P3D)].
OS Cocksackievirus B3 (strain Nancy).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=103903;
RN [1]
RN [2]
RN [3]
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=90204667; PubMed=2157045;
RA Klump W.M., Bergmann I., Mueller B.C., Ameis D., Kandolf R.;
RT "Complete nucleotide sequence of infectious Cocksackievirus B3 cDNA:
RT two initial 5' uridine residues are regained during plus-strand RNA
RT synthesis.";
RL J. Virol. 64:1573-1583(1990).
RP SEQUENCE FROM N.A.
RX MEDLINE=87122156; PubMed=3027968;
RA Lindberg A.M., Staalhandske P.O.K., Pettersson U.;
RT "Genome of cocksackievirus B3.";
RL J. Virol. 51:742-746(1984).
RP SEQUENCE OF 1724-2185 FROM N.A.
RX MEDLINE=84292451; PubMed=608796;
RA Staalhandske P.O.K., Lindberg A.M., Pettersson U.;
RT "Replicase gene of cocksackievirus B3.";
RL J. Virol. 51:742-746(1984).
RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 1-851.
RA Muckelbauer J.K., Kremer M., Minor I., Tong L., Zlotnick A.,
RA Johnson J.E., Rossmann M.G.;
RT "Structure determination of cocksackievirus B3 to 3.5-A resolution.";
RL Acta Crystallogr. D 51:871-887(1995).
CC -!- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
CC AT CERTAIN Q/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE
CC PROTEASES.
CC -!- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
CC MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.
```

```
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/P2A IS
CC CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
CC -!- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -!- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M33854; AAA42931.1; --
CC EMBL; K02709; AAA42932.1; --
CC EMBL; M16572; AAA74400.1; --
CC PIR; A26354; GNNYB3.
CC PIR; A34664; GNNYB3.
CC PDB; 1COV; 08-MAR-96.
CC MEROPS; C03.011; --
CC MEROPS; C03.022; --
CC InterPro: IPR000199; Cys-protease-3C.
CC InterPro: IPR003138; Pico_PIA.
CC InterPro: IPR000081; Pico_P2A.
CC InterPro: IPR002527; Pico_P2B.
CC InterPro: IPR006005; RNA_helicase.
CC InterPro: IPR001205; RNA_pol_P3D.
CC InterPro: IPR001676; Rnv.
CC Pfam; PF00548; Cys-protease-3C; 1.
CC Pfam; PF02226; Pico_PIA; 1.
CC Pfam; PF00947; Pico_P2A; 1.
CC Pfam; PF01552; Pico_P2B; 1.
CC Pfam; PF00073; rhv; 3.
CC Pfam; PF00680; RNA_dep_RNA_pol; 1.
CC Pfam; PF00910; RNA_helicase; 1.
CC ProDom; PD001125; Cys-protease-3C; 1.
CC ProDom; PD001274; Pico_P2B; 1.
CC ProDom; PD001306; Pico_P2A; 1.
CC Polyprotein; Coat protein; Core protein; Core protein; Transferase; Myristate;
CC RNA-directed RNA polymerase; Hydrolase; Thiol protease; 3D-structure.
KW CHAIN 2 69 COAT PROTEIN VP4.
FT CHAIN 70 332 COAT PROTEIN VP2.
FT CHAIN 333 570 COAT PROTEIN VP3.
FT CHAIN 571 851 COAT PROTEIN VP1.
FT CHAIN 852 1001 PICORNAIN 2A.
FT CHAIN 1002 1100 CORE PROTEIN P2B.
FT CHAIN 1101 1429 CORE PROTEIN P2C.
FT CHAIN 1430 1518 CORE PROTEIN P3A.
FT CHAIN 1519 1540 GENOME-LINKED PROTEIN VPG.
FT CHAIN 1541 1723 PICORNAIN 3C.
FT CHAIN 1724 2185 RNA-DIRECTED RNA POLYMERASE.
FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
FT ACT_SITE 1687 1687 PROTEASE (POTENTIAL).
FT ACT_SITE 1701 1701 PROTEASE (POTENTIAL).
FT CONFLICT 16 16 R -> G (IN REF. 2).
FT CONFLICT 177 177 V -> D (IN REF. 2).
FT CONFLICT 469 469 P -> L (IN REF. 2).
FT CONFLICT 487 487 I -> V (IN REF. 2).
FT CONFLICT 510 510 F -> Y (IN REF. 2).
FT CONFLICT 516 516 Y -> C (IN REF. 2).
FT CONFLICT 566 566 Q -> E (IN REF. 2).
FT CONFLICT 593 593 T -> N (IN REF. 2).
FT CONFLICT 650 650 K -> E (IN REF. 2).
FT CONFLICT 854 865 FGQSGAVYVGN -> IWTTRGSVCGD (IN REF. 2).
FT CONFLICT 873 873 L -> S (IN REF. 2).
FT CONFLICT 1097 1097 A -> P (IN REF. 2).
FT CONFLICT 1280 1280 Q -> H (IN REF. 2).
FT CONFLICT 1437 1437 I -> F (IN REF. 2).
FT CONFLICT 1503 1503 V -> M (IN REF. 2).
```

```
FT CONFLICT 1616 1616 K -> E (IN REF. 2).
FT CONFLICT 1624 1624 R -> G (IN REF. 2).
FT CONFLICT 1627 1627 R -> G (IN REF. 2).
FT CONFLICT 1630 1630 L -> V (IN REF. 2).
FT CONFLICT 1718 1718 Y -> N (IN REF. 2).
FT CONFLICT 1734 1734 D -> V (IN REF. 2).
FT CONFLICT 1758 1758 E -> V (IN REF. 2 AND 3).
FT CONFLICT 1824 1824 V -> R (IN REF. 2 AND 3).
FT CONFLICT 1867 1867 C -> R (IN REF. 2 AND 3).
FT CONFLICT 1880 1880 Y -> H (IN REF. 2 AND 3).
FT CONFLICT 2001 2001 D -> N (IN REF. 2 AND 3).
FT CONFLICT 2095 2095 A -> V (IN REF. 2 AND 3).
FT CONFLICT 2115 2115 V -> T (IN REF. 2 AND 3).
FT CONFLICT 2175 2175 S -> T (IN REF. 2 AND 3).
FT CONFLICT 2178 2178 R -> G (IN REF. 2 AND 3).
SQ SEQUENCE 2185 AA; 243451 MW; 1B5ECE3DA47338FF CRC64;
```

```
alignment_scores:
  Quality: 63.00 Length: 26
  Ratio: 3.316 Gaps: 0
  Percent Similarity: 73.077 Percent Identity: 42.308
```

alignment_block:

US-09-049-696-4/rev x POLG_CXB3N

Align seg 1/1 to: POLG_CXB3N from: 1 to: 2185

```
126 AATTCAACTATAGATACACATCTTGTGCAACATTATAGAAGCCTTCTC 77
|||||
89 AsnSerThrIleThrGlnGluCysAlaAsnValValGlyTyrGI 105
```

76 CGTCTCGCGGGATTGGAGACAACACTCA 49

105 yValTrpProAspTyrLeuLysAspSer 114

seq_name: SwissProt_39:POLG_CXB3N

seq_documentation_block:

ID POLG_CXB3N STANDARD; PRT; 2185 AA.

AC Q66282;

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEIN VP4 (PIA); COAT PROTEIN VP2

DE (PIB); COAT PROTEIN VP3 (PIC); COAT PROTEIN VP1 (PID); PICORNAIN 2A

DE (EC 3.4.22.29) (P2A); CORE PROTEIN P2B; CORE PROTEIN P2C; CORE PROTEIN

DE (P3A); GENOME-LINKED PROTEIN VPG (P3B); PICORNAIN 3C (EC 3.4.22.28)

DE (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (P3D)].

OS Cocksackievirus B3 (strain Woodruff).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;

OC Enterovirus.

OX NCBI_TaxID=103904;

RN [1]

RP SEQUENCE FROM N.A.

RA Knowlton K.U., Jeon E.S., Berkley R.W., Wessely R., Huber S.;

RT "A mutation in the puff region of VP2 attenuates the myocarditic

phenotype of an infectious cDNA of the Woodruff virus.";

RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.

CC !- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE

AT CERTAIN O/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE

PROTEASES.

CC !- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR

MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.

CC !- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,

EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,

VP3, AND VP4.

CC !- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/P2A IS

CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.

CC !- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC !- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

EMBL; U57056; AAB02228.1; .

DR MEROPS; C03.011; .

DR MEROPS; C03.022; .

DR InterPro; IPR001643; Calici_coat.

DR InterPro; IPR000199; Cys-protease-3C.

DR InterPro; IPR003138; Pico_P1A.

DR InterPro; IPR000081; Pico_P2A.

DR InterPro; IPR002527; Pico_P2B.

DR InterPro; IPR006055; RNA_helicase.

DR InterPro; IPR001205; RNA_pol_P3D.

DR InterPro; IPR001676; Rnv.

DR Pfam; PF00548; Cys-protease-3C; 1.

DR Pfam; PF02226; Pico_P1A; 1.

DR Pfam; PF00947; Pico_P2A; 1.

DR Pfam; PF01552; Pico_P2B; 1.

DR Pfam; PF00073; rhv; 3.

DR Pfam; PF00680; RNA_dep_RNA_pol; 1.

DR Pfam; PF00910; RNA_helicase; 1.

DR PRINTS; PR00918; CALICIVIRUSNS.

DR ProDom; PD001125; Cys-protease-3C; 1.

DR ProDom; PD001274; Pico_P2B; 1.

DR ProDom; PD001306; Pico_P2A; 1.

KW Polyprotein; Coat protein; Core protein; Transferase; Myristate;

KW RNA-directed RNA polymerase; Hydrolase; Thiol protease.

FT CHAIN 2 69

FT CHAIN 70 332

FT CHAIN 333 570

FT CHAIN 571 851

FT CHAIN 852 1001

FT CHAIN 1002 1100

FT CHAIN 1101 1429

FT CHAIN 1430 1518

FT CHAIN 1519 1540

FT CHAIN 1541 1723

FT CHAIN 1724 2185

FT LIPID 2 2

FT ACT_SITE 1687 1687

FT ACT_SITE 1701 1701

SQ SEQUENCE 2185 AA; 243680 MW; FD93A677904252FA CRC64;

alignment_scores:

Quality: 63.00 Length: 26

Ratio: 3.316 Gaps: 0

Percent Similarity: 73.077 Percent Identity: 42.308

alignment_block:

US-09-049-696-4/rev x POLG_CXB3N

Align seg 1/1 to: POLG_CXB3N from: 1 to: 2185

126 AATTCAACTATAGATACACATCTTGTGCAACATTATAGAAGCCTTCTC 77

|||||

89 AsnSerThrIleThrGlnGluCysAlaAsnValValGlyTyrGI 105

76 CGTCTCGCGGGATTGGAGACAACACTCA 49

105 yValTrpProAspTyrLeuLysAspSer 114

seq_name: SwissProt_39:A33_PLEWA

seq_documentation_block:

ID A33_PLEWA STANDARD; PRT; 624 AA.

AC Q02084;

DT 01-OCT-1994 (Rel. 30, Created)

01-OCT-1994 (Rel. 30, Last sequence update)
 20-AUG-2001 (Rel. 40, Last annotation update)
 ZINC-BINDING PROTEIN A33.
 Pleurodeles waltlii (Iberian ribbed newt).
 Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae;
 Pleurodeles.
 NCBI_TaxID=8319;
 [1]
 SEQUENCE FROM N.A.
 TISSUE-Ovary;
 MEDLINE=93154311; PubMed=7679068;
 Bellini M., Lacroix J.-C., Gall J.G.;
 RT "A putative zinc-binding protein on lampbrush chromosome loops.";
 EMBO J. 12:107-114(1993).
 CC -1- FUNCTION: MAY BE A NUCLEAR REGULATORY PROTEIN THAT IS STORED
 IN THE GERMINAL VESICLE FOR USE DURING EARLY EMBRYOGENESIS
 AND MAY PLAY A ROLE IN THE SYNTHESIS OR PROCESSING OF PRE-MRNA
 DURING OOGENESIS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- DEVELOPMENTAL STAGE: IT FIRST APPEARS ON THE CHROMOSOME LOOPS
 AND IN THE NUCLEOLUS OF THE GERMINAL VESICLE (GV). IT IS
 TRANSMITTED TO THE EGG AT GV BREAKDOWN AND APPEARS IN EMBRYONIC
 NUCLEI AT THE MID-BLASTULA STAGE AND IS FOUND IN MANY BUT NOT
 ALL NUCLEI AT STILL LATER STAGES OF EMBRYOGENESIS.
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC -1- SIMILARITY: CONTAINS 1 B BOX-TYPE ZINC FINGER.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC EMBL; L04190; AAA49614.1;
 InterPro; IPR003649; Bbox_C.
 DR InterPro; IPR003879; Butyroph_DUF_C.
 DR InterPro; IPR000953; Chromo.
 DR InterPro; IPR001870; Gamma_carboxylase.
 DR InterPro; IPR003877; SPRY.
 DR InterPro; IPR000315; Znf_bbox.
 DR InterPro; IPR002991; Znf_bbox_1.
 DR InterPro; IPR001841; Znf_fing.
 DR Pfam; PF00622; SPRY; 1.
 DR Pfam; PF00643; zf-B_box; 1.
 DR Pfam; PF00097; zf-C3HC4; 1.
 DR SMART; SM00502; BBOX; 1.
 DR SMART; SM00336; BBOX; 1.
 DR SMART; SM00298; CHROMO; 1.
 DR SMART; SM00184; RING; 1.
 DR SMART; SM00449; SPRY; 1.
 DR PROSITE; PS00119; ZF_BBOX; 1.
 DR PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
 KW Zinc-finger; Nuclear protein; Developmental protein; Coiled coil;
 RNA-binding.
 FT DOMAIN 19 26 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 134 149 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT ZN_FING 162 201 RING-TYPE.
 FT ZN_FING 238 269 B BOX-TYPE.
 FT DOMAIN 337 386 COILED COIL.
 FT DOMAIN 449 619 RFP-LIKE.
 SQ SEQUENCE 624 AA; 71056 MW; 60DBD1F3F071EFFD CRC64;

alignment_scores:
 Quality: 62.50 Length: 44
 Ratio: 2.500 Gaps: 2
 Percent Similarity: 56.818 Percent Identity: 36.364

alignment_block:

US-09-049-696-4 x A33_PLEWA

Align_seg 1/1 to: A33_PLEWA from: 1 to: 624
 47 TGTGAGTTTCTT.....CTCCAATCCCGCAGACGGA 78
 |||||
 368 CysGluPheilleGluGluAlaIleSerThrGlnSerArgLeuAsnGI 384
 |||||
 79 GAAGGCTTCTATATGTTTCACAAATGTTGATTCTATAGTTGAATCT 128
 |||||
 384 userAspProilleAlaPheLeuThrAspilleLysSerPheilleGluLysC 401
 |||||
 129 GTACAGAACAAACCAACACAAAGAGAGCTCCA 160
 |||||
 401 yscysGluGlu...HisArgLysGlyValPro 410
 |||||

seq_name: SwissProt_39:SRP68_HUMAN

seq_documentation_block:
 ID SRP68_HUMAN STANDARD; PRT; 619 AA.

AC GSUHB9;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SIGNAL RECOGNITION PARTICLE 68 KDA PROTEIN (SRP68).
 GN SRP68.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 MEDLINE=20087201; PubMed=10618370;
 Politz J.C., Varovoi S., Kilroy S.M., Gowda K., Zwiab C., Pederson T.;
 RT "Signal recognition particle components in the nucleolus.";
 proc. Natl. Acad. Sci. U.S.A. 97:55-60(2000).
 CC -1- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE
 IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC
 RETICULUM MEMBRANE. SRP68 BINDS THE 7S RNA. SRP72 BINDS TO THIS
 COMPLEX SUBSEQUENTLY. THIS RIBONUCLEOPROTEIN COMPLEX MIGHT
 INTERACT DIRECTLY WITH THE DOCKING PROTEIN IN THE ER MEMBRANE AND
 POSSIBLY PARTICIPATE IN THE ELONGATION ARREST FUNCTION.
 CC -1- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE
 OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,
 SRP19, SRP14 AND SRP9.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR; NUCLEOLAR.
 CC -1- MISCELLANEOUS: THE RNA BINDING DOMAIN IS LOCATED NEAR THE
 N-TERMINUS.
 CC -1- SIMILARITY: BELONGS TO THE SRP68 FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC EMBL; AFI95951; AAF24308.1;
 DR MIM; 604858;
 KW Signal recognition particle; Ribonucleoprotein; RNA-binding;
 FT DOMAIN 9 24 POLY-GLY.
 SQ SEQUENCE 619 AA; 70242 MW; A09FA9A7338C36E0 CRC64;

alignment_scores:
 Quality: 62.00 Length: 57
 Ratio: 1.771 Gaps: 2
 Percent Similarity: 61.404 Percent Identity: 29.825

alignment_block:

US-09-049-696-4 x SRP68_HUMAN

Align seg 1/1 to: SR68_HUMAN from: 1 to: 619

```
14 TTCAATAAAGTAACAGAGCTCTATGAAAGAGTGTGAGTTGTCTTCCA 63
   |||||
203 PheAsnLysCysLysThrIleTyrGluLys.....LeuAl 214
   |||||
64 ATCCCGCCAGCAGGAGAGCTCTATAATGTTTGCACACATGTTGATT 113
   |||||
214 aSerAlaPheThrGluGluGlnAlaValLeuTyrAsnGlnArgValGluG 231
   |||||
114 CTATA.....GTTGAATTCGTACAGAACAAACCAACCAAGAA 154
   |||||
231 lulleSerProAsnIleArgTyrCysAlaTyrAsnIleGlyAspGlnSer 247
   |||||
155 GCTCCAAACAGCAAAATCAA 175
   |||||
248 AlaIleAsnGluLeuMetGln 254
```

seq_name: SwissProt_39:SR68_CANFA

seq_documentation_block:

ID SR68_CANFA STANDARD; PRT; 622 AA.

AC 000004;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE SIGNAL RECOGNITION PARTICLE 68 KDA PROTEIN (SRP68).

GN SRP68.

OS Canis familiaris (Dog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

OX NCBI_TaxID=9615;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE=91092392; PubMed=17023390;

RA Herz J., Flint N., Stanley K., Frank R., Dobberstein B.;

RT "The 68 kda protein of signal recognition particle contains a

glycine-rich region also found in certain RNA-binding proteins.";

RL FEBS Lett. 276:103-107(1990).

CC -!- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY HAS A CRUCIAL ROLE

IN TARGETING SECRETORY PROTEINS TO THE ROUGH ENDOPLASMIC

RETICULUM MEMBRANE. SRP68 BINDS THE 7S RNA. SRP72 BINDS TO THIS

COMPLEX SUBSEQUENTLY. THIS RIBONUCLEOPROTEIN COMPLEX MIGHT

INTERACT DIRECTLY WITH THE DOCKING PROTEIN IN THE ER MEMBRANE AND

POSSIBLY PARTICIPATE IN THE ELONGATION ARREST FUNCTION.

CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE

OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,

SRP19, SRP14 AND SRP9.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR; NUCLEOLAR (BY

SIMILARITY).

CC -!- MISCELLANEOUS: THE RNA BINDING DOMAIN IS LOCATED NEAR THE

N-TERMINUS.

CC -!- SIMILARITY: BELONGS TO THE SRP68 FAMILY.

CC -!- CAUTION: SOME AUTHORS FOUND GENOMIC CLONES THAT HAVE 9 OR 12

CONSECUTIVE GLYCINE RESIDUES INSTEAD OF 15 (AA 9-27).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X53744; CAA37773.1; ALT_SEQ.

DR PIR; S12981; s12981.

KW Signal recognition particle; Ribonucleoprotein; RNA-binding;

Nuclear protein.

FT DOMAIN 9 27 POLY-GLY.

SQ SEQUENCE 622 AA; 70275 MW; DBQ3DFB0DAEB942 CRC64;

alignment_scores:

```
Quality: 62.00 Length: 57
Ratio: 1.771 Gaps: 2
Percent Similarity: 61.404 Percent Identity: 29.825

alignment_block:
US-09-049-696-4 x SR68_CANFA
..
Align seg 1/1 to: SR68_CANFA from: 1 to: 622
..
14 TTCAATAAAGTAACAGAGCTCTATGAAAGAGTGTGAGTTGTCTTCCA 63
   |||||
208 PheAsnLysCysLysThrIleTyrGluLys.....LeuAl 219
   |||||
64 ATCCCGCCAGCAGGAGAGCTCTATAATGTTTGCACACATGTTGATT 113
   |||||
219 aSerAlaPheThrGluGluGlnAlaValLeuTyrAsnGlnArgValGluG 236
   |||||
114 CTATA.....GTTGAATTCGTACAGAACAAACCAACCAAGAA 154
   |||||
236 lulleSerProAsnIleArgTyrCysAlaTyrAsnIleGlyAspGlnSer 252
   |||||
155 GCTCCAAACAGCAAAATCAA 175
   |||||
258 AlaIleAsnGluLeuMetGln 259
   |||||
seq_name: SwissProt_39:O6B1_HUMAN
seq_documentation_block:
ID O6B1_HUMAN STANDARD; PRT; 311 AA.
AC O95007;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE OLFATORY RECEPTOR 6B1 (OLFATORY RECEPTOR 7-3) (OR7-3).
GN O6B1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Baher C., Williams D.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```

FT DOMAIN 220 234 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 235 257 6 (POTENTIAL).
FT DOMAIN 258 270 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 271 290 7 (POTENTIAL).
FT DOMAIN 291 311 CYTOPLASMIC (POTENTIAL).
FT DISULFID 97 189 BY SIMILARITY.
FT CARBOHYD 5 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 311 AA; 35299 MW; 671FD19658FE0616 CRC64;

alignment_scores:
  Quality: 61.00 Length: 77
  Ratio: 1.452 Gaps: 6
Percent Similarity: 54.545 Percent Identity: 29.870

alignment_block:
US-09-049-696-4/rev x 06BL_HUMAN ..
Align seg 1/1 to: 06BL_HUMAN from: 1 to: 311

181 GCATTTTTCATTTGCTGCTTTGGAGCTCTTGTGTGGTTTGTCTG 132
   ||||| :||| :||| :||| :||| :||| :||| :||| :|||
93 SerPheThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 109
   ||||| :||| :||| :||| :||| :||| :||| :||| :|||
131 TACAGAA.....TTCAACTAT...AGAATCAACATGT 103
   ||||| :||| :||| :||| :||| :||| :||| :||| :|||
109 sThrGluCysValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 126
   ||||| :||| :||| :||| :||| :||| :||| :||| :|||
102 TGTGCAAA.....CATTAGAGCTTCTC..... 77
   ||||| :||| :||| :||| :||| :||| :||| :||| :|||
126 leCysArgProLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 142
   ||||| :||| :||| :||| :||| :||| :||| :||| :|||
76 CGTCTGCGGGATTTGAGCAACAACTACATCTTTTCATAGAG..... 32
   ||||| :||| :||| :||| :||| :||| :||| :||| :|||
143 ArgLeuAla...LeuGlySerTrpAlaIleGlyPheGlyLeuAla 158
   ||||| :||| :||| :||| :||| :||| :||| :||| :|||
31 .TCCTGTACTTATTCATGATGCATCTTT 2
   ||||| :||| :||| :||| :||| :||| :||| :||| :|||
158 aLysIleTyrPheIleSerCysLeuSerPhe 168

seq_name: SwissProt_39:INX6_DROME

seq_documentation_block:
ID INX6_DROME STANDARD; PRT; 481 AA.
AC Q9VR82; Q9UAI5;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INX6INX6 (INX6INX6-6) (GAP JUNCTION PROTEIN PRP6) (PAS-RELATED
  PROTEIN 6)
GN INX6 OR PRP6 OR CG2977.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Phyllophaga; Drosophilidae; Drosophilinae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.F., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.G., Rogers J.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berland B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

```

```

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Gladek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Haris N.D., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulyov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT The genome sequence of Drosophila melanogaster.;
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE OF 228-431 FROM N.A.
RC TISSUE=Head;
RX MEDLINE=99282524; PubMed=10352230;
RA Curtin K.D., Zhang Z., Wyman R.J.;
RA Drosophila has several genes for gap junction proteins.;
RL Gene 232:191-201(1999).
CC -!- FUNCTION: STRUCTURAL COMPONENTS OF THE GAP JUNCTIONS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE INX6INX6 FAMILY.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
  FRAMESHIFT IN POSITION 240.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE003572; AAF50922.1;
DR EMBL; AF137271; AD50380.1; ALT_FRAME.
DR FlyBase; FBgn0027107; inx6.
DR InterPro; IPR000990; Innexin.
DR Pfam; PF00876; Innexin; 2.
DR PRINTS; PR01262; INNEXIN.
DR Gap junction; Transmembrane.
KW DOMAIN 1 21 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 22 42 POTENTIAL.
FT DOMAIN 43 144 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 145 165 POTENTIAL.
FT DOMAIN 166 220 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 221 241 POTENTIAL.
FT DOMAIN 242 302 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 303 323 POTENTIAL.
FT DOMAIN 324 481 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 418 431 S -> T (IN REF. 2).
FT CONFLICT 431 431 S -> R (IN REF. 2).
SQ SEQUENCE 481 AA; 55661 MW; A435A51566B83013 CRC64;

alignment_scores:
  Quality: 61.00 Length: 72
  Ratio: 1.848 Gaps: 2
Percent Similarity: 45.833 Percent Identity: 23.611

alignment_block:

```


US-09-049-696-4/rev x INX6_DROME

Align seg 1/1 to: INX6_DROME from: 1 to: 481

178 TTTTGTGTTTGGTGTGTTGGAGCTCTTCTGTTGTGTTGTTGTTCTGTAC 129
 304 PheAlaValLeuTyValTrpPheLeuPheLeAlaLeuLeuAlaIleMe 320
 128 AGAATCACTATAGATCAACATCTTGTGCG..... 98
 320 tAsnIleLeuTyArgLeuLeuValIleCysCysProGluLeuArgLeuG 337
 97NAACATTATAGAGC 83
 337 lnLeuLeuArgTrpHisLeuAsnGlyMetProLysSerHisValArgGlu 353
 82 CTTCCTCGTCTGGCGGATTTGGAGCAAACTCACATCTTTTTCATAGA 33
 354 ValLeuAlaSerAlaGlyTyArgGlyAsp.....TrpPheValLe 366
 32 GTCTCTGTTACTTTATT 17
 366 uMetCysValSerIle 371

seq_name: SwissProt_39:DPOL_HPBHE

seq_documentation_block:
 ID DPOL_HPBHE STANDARD; PRT: 788 AA.

AC P13846;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE P PROTEIN (INCLUDES: DNA-DIRECTED DNA POLYMERASE (EC 2.7.7.7); RNA-DIRECTED DNA POLYMERASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)).
 GN P.
 OS Heron hepatitis b virus.
 OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
 OX NCBI_TaxID=28300;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8833160; PubMed=3418788;
 RA Sprengel R., Kaleta E.F., Will H.;
 RT "Isolation and characterization of a hepatitis B virus endemic in herons";
 RL J. Virol. 62:3832-3839(1988).
 CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE + DNA(N).
 CC -!- CATALYTIC ACTIVITY: ENDONUCLEOLYTIC CLEAVAGE TO 5'-PHOSPHOMONESTER.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M22056; AAA45738.1;
 DR PIR; A30082; JDVLHH.
 DR InterPro; IPR001462; DNAPol_viral_C.
 DR InterPro; IPR000201; DNAPol_viral_N.
 DR InterPro; IPR000477; RVTse.
 DR Pfam; PF00336; DNAPol_viral_C.1.
 DR Pfam; PF00242; DNAPol_viral_N.1.
 DR Pfam; PF00078; rvt; 1.
 DR ProDom; PD000814; DNAPol_viral_C.1.
 KW Transferase; RNA-directed DNA polymerase; DNA-directed DNA polymerase;
 KW Hydrolyase; Nuclease; Endonuclease; DNA replication; DNA-binding.
 SQ SEQUENCE 788 AA; 90070 MW; FB44F38F75EADF44 CRC64;

alignment_scores:

Quality: 61.00 Length: 55
 Ratio: 1.743 Gaps: 2
 Percent Similarity: 63.636 Percent Identity: 32.727

alignment_block:

US-09-049-696-4 x DPOL_HPBHE

Align seg 1/1 to: DPOL_HPBHE from: 1 to: 788

AGATGCACATTCAATAAGTAACAGGACTCTATGAA...AAAGGATGTGA 51
 -||||||.....|||.....|||.....|||.....|||.....
 86 ArgHisLysLeuGlyLysLeuThrGlyLeuTyrglnMetLysGlyCysG 102
 52 GTTGTCTTCAATCCCGCAGGAGGAGGCTTCTATATGTTGTCAC 101
 -||||||.....|||.....|||.....|||.....|||.....
 102 uPheAsnProHisTrpLysIleProAspIleSerAlaThrAsnPheSerG 119
 102 AACATGTTGATTCATAGTTGAATTCGTACAGCAACAAACCAACAAA 151
 -||||||.....|||.....|||.....|||.....|||.....
 119 ln.....GluIleIleasnGluCysProSerArgAsnTrpLysTr 132
 152 GAAGCTCCAAACAAG 166
 -||||||.....
 133 LeuThrProAlaLys 137

seq_name: SwissProt_39:POLG_CXB4E

seq_documentation_block:
 ID POLG_CXB4E STANDARD; PRT: 2183 AA.

AC Q86887;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE GENOME POLYPROTEIN (CONTAINS: COAT PROTEIN VP4 (PIA); COAT PROTEIN VP2 (PIB); COAT PROTEIN VP3 (PIC); COAT PROTEIN VP1 (PID); PICORNAIN 2A (EC 3.4.22.29) (P2A); CORE PROTEIN P2B; CORE PROTEIN P2C; CORE PROTEIN P3A; GENOME-LINKED PROTEIN VP3 (P3B); PICORNAIN 3C (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (P3D)).
 OS Coxsackievirus B4 (strain E2).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OX Enterovirus.
 OX NCBI_TaxID=103905;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=95205102; PubMed=7897366;
 RA Kang Y., Chatterjee N.K., Nodwell M.J., Yoon J.W.;
 RT "Complete nucleotide sequence of a strain of coxsackie B4 virus of human origin that induces diabetes in mice and its comparison with nondiabetogenic coxsackie B4 JBV strain";
 RL J. Med. Virol. 44:353-361(1994).
 CC -!- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE AT CERTAIN Q/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE PROTEASES.

CC -!- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.
 CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.
 CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS. CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC. VP1/P2A IS CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
 CC -!- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -!- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; S76772; AAB33885.1; -

```
DR InterPro: IPR001643; Calici.coat.
DR InterPro: IPR000199; Cys-protease-3C.
DR InterPro: IPR000318; Pico_P1A.
DR InterPro: IPR000081; Pico_P2A.
DR InterPro: IPR0002527; Pico_P2B.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR InterPro: IPR001676; Rnv.
DR Pfam: PF00548; Cys-protease-3C; 1.
DR Pfam: PF02226; Pico_P1A; 1.
DR Pfam: PF00947; Pico_P2A; 1.
DR Pfam: PF01552; Pico_P2B; 1.
DR Pfam: PF00073; rhv; 3.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
DR PRINTS: PK00918; CALICIVIRUS.
DR ProDom: PD001125; Cys-protease-3C; 1.
DR ProDom: PD001274; Pico_P2B; 1.
DR ProDom: PD001306; Pico_P2A; 1.
DR PolyProtein: Coat protein; Core protein; Hydrolase; Thiol protease.
DR RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 2 69
FT CHAIN 70 330
FT CHAIN 331 568
FT CHAIN 569 849
FT CHAIN 850 999
FT CHAIN 1000 1098
FT CHAIN 1099 1427
FT CHAIN 1428 1516
FT CHAIN 1517 1538
FT CHAIN 1539 1721
FT CHAIN 1722 2183
FT LIPID 2
FT ACT_SITE 1685 1685
FT ACT_SITE 1699 1699
FT ACT_SITE 1699 1699
FT ACT_SITE 2183 AA; 244649 MW; BF44781CF6981D39 CRC64;
SQ SEQUENCE 2183 AA; 244649 MW; BF44781CF6981D39 CRC64;

alignment_scores:
  Quality: 61.00 Length: 22
  Ratio: 3.588 Gaps: 0
Percent Similarity: 77.273 Percent Identity: 45.455

alignment_block:
US-09-049-696-4/rev x POLG_CXB4E ..
Align seg 1/1 to: POLG_CXB4E from: 1 to: 2183

126 AATTCACTATAGATCANCATGTTGTGCAACATTATAGAGCTTCTC 77
|||||
89 AsnSerThrIleThrThrGlnGluCysAlaAsnValValValGlyIyrGI 105
|||||

76 CGTCTGGCGGATTCG 61
|||||
105 yValTrpAsnIyr 110
|||||

seq_name: SwissProt_39:POLG_CXB4J
seq_documentation_block:
ID POLG_CXB4J STANDARD; PRT; 2183 AA.
AC P08292;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEIN VP4 (P1A); COAT PROTEIN VP2
DE (P1B); COAT PROTEIN VP3 (P1C); COAT PROTEIN VP1 (PID); PICORNAIN 2A
DE (EC 3.4.22.29) (P2A); CORE PROTEIN P2B; CORE PROTEIN P2C; CORE PROTEIN
DE P3A; GENOME-LINKED PROTEIN VPG (P3B); PICORNAIN 3C (EC 3.4.22.28)
DE (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (P3D)].
OS Cocksackievirus B4 (strain JVB / Benschoten / New York/51)
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OC NCBI_TaxID=103906;
```

```
KN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87253111; PubMed=3037008;
RA Jenkins O., Booth J.D., Minor P.D., Almond J.W.;
RT "The complete nucleotide sequence of coxsackievirus B4 and its
RL J. Gen. Virol. 68:1835-1848(1987).
RN [2]
RP REVISIONS.
RA Jenkins O.;
RL Submitted (AUG-1989) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
CC AT CERTAIN Q/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE
CC PROTEASES.
CC -!- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
CC MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/VP2 IS
CC CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
CC -!- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -!- SIMILARITY: P2C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05690; CAA29172.1; -.
DR PIR; A27170; GNMYB4.
DR HSP; P03313; LCOV.
DR MEROPS; C03.011; -.
DR MEROPS; C03.022; -.
DR InterPro: IPR000199; Cys-protease-3C.
DR InterPro: IPR000318; Pico_P1A.
DR InterPro: IPR000081; Pico_P2A.
DR InterPro: IPR0002527; Pico_P2B.
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR InterPro: IPR001676; Rnv.
DR Pfam: PF00548; Cys-protease-3C; 1.
DR Pfam: PF02226; Pico_P1A; 1.
DR Pfam: PF00947; Pico_P2A; 1.
DR Pfam: PF01552; Pico_P2B; 1.
DR Pfam: PF00073; rhv; 3.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
DR ProDom: PD001125; Cys-protease-3C; 1.
DR ProDom: PD001274; Pico_P2B; 1.
DR ProDom: PD001306; Pico_P2A; 1.
DR PolyProtein; Coat protein; Core protein; Hydrolase; Thiol protease.
DR RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 2 69
FT CHAIN 70 330
FT CHAIN 331 568
FT CHAIN 569 849
FT CHAIN 850 999
FT CHAIN 1000 1098
FT CHAIN 1099 1427
FT CHAIN 1428 1516
FT CHAIN 1517 1538
FT CHAIN 1539 1721
FT CHAIN 1722 2183
FT LIPID 2
FT ACT_SITE 1685 1685
FT ACT_SITE 1699 1699
FT ACT_SITE 1699 1699
FT ACT_SITE 2183 AA; 244012 MW; DE069DE3AE91AF0 CRC64;
SQ SEQUENCE 2183 AA; 244012 MW; DE069DE3AE91AF0 CRC64;
```

alignment_scores:
 Quality: 61.00 Length: 22
 Ratio: 3.588 Gaps: 0
 Percent Similarity: 77.273 Percent Identity: 45.455

alignment_block:
 US-09-049-696-4/rev x POLG_CXB4J ..
 Align seg 1/1 to: POLG_CXB4J from: 1 to: 2183

126 AATCACTACTAGATCAACATGTTGCAACACATTATAGACGCTCTC 77
 |||||
 89 AsnSerThrIleThrGlnGluCysAlaAsnValValGlyTyrGI 105
 |||||

76 CGTCTGCGGGGATTGG 61
 :|||
 105 yValTrpProAspTyr 110

seq_name: SwissProt_39:FLIM_BUCAI

seq_documentation_block:
 ID FLIM_BUCAI STANDARD; PRT; 315 AA.
 AC P57182;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE FLAGELLAR MOTOR SWITCH PROTEIN FLIM.
 GN FLIM OR BU080.

OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum)
 OS symbiotic bacterium)
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 OX NCBI_TaxID=118099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TKYO 1998;
 RX MEDLINE=20445173; PubMed=10993077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 Buchnera sp. APS.";
 RL Nature 407:81-86(2000).
 CC -!- FUNCTION: FLIM IS ONE OF THREE PROTEINS (FLIM, FLIN, FLIM) THAT
 CC FORM A SWITCH COMPLEX THAT IS PROPOSED TO BE LOCATED AT THE BASE
 CC OF THE BASAL BODY. THIS COMPLEX INTERACTS WITH THE CHEY AND CHEZ
 CC CHEMOTAXIS PROTEINS, IN ADDITION TO CONTACTING COMPONENTS OF THE
 CC MOTOR THAT DETERMINE THE DIRECTION OF FLAGELLAR ROTATION (BY
 CC SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE FLIM FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.

DR EMBL; AP001118; BAB12800.1;
 KW Chemotaxis; Flagella; Flagellar rotation; Membrane; Complete proteome.
 SQ SEQUENCE 315 AA; 36939 MW; FE5CC11D9573F198 CRC64;

alignment_scores:
 Quality: 59.50 Length: 50
 Ratio: 1.700 Gaps: 1
 Percent Similarity: 70.000 Percent Identity: 26.000

alignment_block:
 US-09-049-696-4 x FLIM_BUCAI ..
 Align seg 1/1 to: FLIM_BUCAI from: 1 to: 315

17 AATAAGTAAACAGAGCTCTATGAAAAAGAGTGTGAGTTTGTCTTCAATC 66
 |||||
 263 AsnLysIleThrGlyPheIleGlnAspGlnAlaIlePheLeuGlyAsnTy 285
 |||||
 67 CGCCACACGAGAGGCTTCTATAATGTTTGCACACATGTTCTATCTA 116
 |||||
 285 rLysArgPheAsnGlnSerIleIlePhe.....IleGluGluP 299
 |||||
 117 TAGTTGAATTCTGTACAGAACAAACACACAAAGAGCTCCAAACAAG 166
 |||||
 299 helieAspSerSerSerGluSerAsnGlnAspLysGluTyrSerAsnGlu 315

seq_name: SwissProt_39:A2A2_RAT

seq_documentation_block:
 ID A2A2_RAT STANDARD; PRT; 937 AA.
 AC P18484;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE ADAPTOR-RELATED PROTEIN COMPLEX 2 ALPHA 2 SUBUNIT (ALPHA-ADAPTIN C)
 DE (CLATHRIN ASSEMBLY PROTEIN COMPLEX 2 ALPHA-C LARGE CHAIN) (100 KDA
 DE COATED VESICLE PROTEIN C) (PLASMA MEMBRANE ADAPTOR HA2/AP2 ADAPTIN
 DE ALPHA C SUBUNIT).
 GN AP2A2 OR ADTAB.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90384857; PubMed=2402467;
 RA Tucker K.L., Nathanson K., Kirchhausen T.;
 RT "Sequence of the rat alpha c large chain of the clathrin associated
 RT protein complex Ap-2.";
 RL Nucleic Acids Res. 18:5306-5306(1990).
 RN [2]
 RP SEQUENCE OF 1-17.

CC TISSUE=Brain;
 CC MEDLINE=89202379; PubMed=2495531;
 RA Kirchhausen T., Nathanson K.L., Matsui W., Valsberg A., Chow E.P.,
 RA Burne C., Keen J.H., Davis A.E.;
 RT "Structural and functional division into two domains of the large
 RT (100- to 115-kDa) chains of the clathrin-associated protein complex
 RT AP-2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2612-2616(1989).
 CC -!- FUNCTION: ADAPTINS ARE COMPONENTS OF THE ADAPTOR COMPLEXES
 CC WHICH LINK CLATHRIN TO RECEPTORS IN COATED VESICLES. CLATHRIN-
 CC ASSOCIATED PROTEIN COMPLEXES ARE BELIEVED TO INTERACT WITH THE
 CC CYTOPLASMIC TAILS OF MEMBRANE PROTEINS, LEADING TO THEIR SELECTION
 CC AND CONCENTRATION. ALPHA ADAPTIN IS A SUBUNIT OF THE PLASMA
 CC MEMBRANE ADAPTOR.
 CC -!- SUBUNIT: ASSEMBLY PROTEIN COMPLEX 2 (AP-2) IS A HETEROTETRAMER
 CC COMPOSED OF TWO LARGE CHAINS (ALPHA AND BETA), A MEDIUM CHAIN
 CC (AP50) AND A SMALL CHAIN (AP17).
 CC -!- SUBCELLULAR LOCATION: COMPONENT OF THE COAT SURROUNDING THE
 CC CYTOPLASMIC FACE OF COATED VESICLES IN THE PLASMA MEMBRANE.
 CC -!- SIMILARITY: BELONGS TO THE ADAPTOR COMPLEXES LARGE SUBUNITS
 CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.

DR EMBL; X53773; CAA37791.1;
 DR PIR; S11276; S11276.
 DR InterPro; IPR002553; Adaptin_N.
 DR InterPro; IPR003164; Alpha_adaptin_C.
 DR Pfam; PF01602; Adaptin_N; 1.

DR pfam; PF02296; Alpha_adaptin_C; 1.
KW Coated pits.
FT INIT_MET
SQ SEQUENCE 937 AA; 103913 MW; 6E62837AB58090CC CRC64;

alignment_scores:
Quality: 59.00 Length: 39
Ratio: 2.269 Gaps: 1
Percent Similarity: 66.667 Percent Identity: 41.026

alignment_block:
US-09-049-696-4 x A2A2_RAT ..

Align seg 1/1 to: A2A2_RAT from: 1 to: 937

63 ATCCCGCCGACGAGAGGCTTCTATATGTTGCCACAAATGTTGAT 112
:::||||| ||| ||||| |||||:||||: 285
271 AspProAlaValArgGlyLeuThrGluCysLeuGluThrIle..... 285
113 TCTATAGTTGAATTCGTACAGACAAACCAACCAAGAGCTCCAAA 162
286LeuAsnLysAlaGlnGluProProlLysSerLysValGlnH 300

163 CAAGCAAAATCAAAAAT 179
||||| |||||
300 isSerAsnAlaLysAsn 305

seq_name: SwissProt_39:A2A2_MOUSE

seq_documentation_block:

ID A2A2_MOUSE STANDARD; PRT; 938 AA.
AC P17427;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE ADAPTOR-RELATED PROTEIN COMPLEX 2 ALPHA-2 SUBUNIT (ALPHA-ADAPTIN C)
DE (CLATHRIN ASSEMBLY PROTEIN COMPLEX 2 ALPHA-C LARGE CHAIN) (100 KDA
DE COATED VESICLE PROTEIN C) (PLASMA MEMBRANE ADAPTOR HAZ/AP2 ADAPTIN
DE ALPHA C SUBUNIT).
GN AP2A2 OR ADTAB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=89155572; PubMed=2564002;
RA Robinson M.S.;
RT "Cloning of cDNAs encoding two related 100-kD coated vesicle proteins
(alpha-adaptins).";
RL J. Cell Biol. 108:833-842(1989).
CC -!- FUNCTION: ADAPTINS ARE COMPONENTS OF THE ADAPTOR COMPLEXES
CC WHICH LINK CLATHRIN TO RECEPTORS IN COATED VESICLES. CLATHRIN-
CC ASSOCIATED PROTEIN COMPLEXES ARE BELIEVED TO INTERACT WITH THE
CC CYTOPLASMIC TAILS OF MEMBRANE PROTEINS, LEADING TO THEIR SELECTION
CC AND CONCENTRATION. ALPHA ADAPTIN IS A SUBUNIT OF THE PLASMA
CC MEMBRANE ADAPTOR.
CC -!- SUBUNIT: ASSEMBLY PROTEIN COMPLEX 2 (AP-2) IS A HETEROTETRAMER
CC COMPOSED OF TWO LARGE CHAINS (ALPHA AND BETA), A MEDIUM CHAIN
CC (AP50) AND A SMALL CHAIN (AP17).
CC -!- SUBCELLULAR LOCATION: COMPONENT OF THE COAT SURROUNDING THE
CC CYTOPLASMIC FACE OF COATED VESICLES IN THE PLASMA MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE ADAPTOR COMPLEXES LARGE SUBUNITS
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.

CC EMBL; X14972; CAA33097.1;
CC PIR; B30111; B30111.
CC FIC; S12471; S12471.
CC MGI; MGI:101920; Ap2a2.
CC InterPro; IPR002553; Adaptin_N.
CC InterPro; IPR003164; Alpha_adaptin_C.
CC Pfam; PF01602; Adaptin_N; 1.
CC Pfam; PF02296; Alpha_adaptin_C; 1.
KW Coated pits.
SQ SEQUENCE 938 AA; 104100 MW; 10C92E0C7AE9DBC CRC64;

alignment_scores:
Quality: 59.00 Length: 39
Ratio: 2.269 Gaps: 1
Percent Similarity: 66.667 Percent Identity: 41.026

alignment_block:

US-09-049-696-4 x A2A2_MOUSE ..

Align seg 1/1 to: A2A2_MOUSE from: 1 to: 938

63 ATCCCGCCGACGAGAGGCTTCTATATGTTGCCACAAATGTTGAT 112
:::||||| ||| ||||| |||||:||||: 286
272 AspProAlaValArgGlyLeuThrGluCysLeuGluThrIle..... 286
113 TCTATAGTTGAATTCGTACAGACAAACCAACCAAGAGCTCCAAA 162
287LeuAsnLysAlaGlnGluProProlLysSerLysValGlnH 301

163 CAAGCAAAATCAAAAAT 179
||||| |||||
301 isSerAsnAlaLysAsn 306

seq_name: SwissProt_39:A2A2_HUMAN

seq_documentation_block:

ID A2A2_HUMAN STANDARD; PRT; 939 AA.
AC O94973; O75403;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ADAPTER-RELATED PROTEIN COMPLEX 2 ALPHA-2 SUBUNIT (ALPHA-ADAPTIN C)
DE (CLATHRIN ASSEMBLY PROTEIN COMPLEX 2 ALPHA-C LARGE CHAIN) (100 KDA
DE COATED VESICLE PROTEIN C) (PLASMA MEMBRANE ADAPTOR HAZ/AP2 ADAPTIN
DE ALPHA C SUBUNIT) (HUNTINGTIN-INTERACTING PROTEIN HYPJ).
GN AP2A2 OR ADTAB OR KIAA0899.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 11-939 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RA "Prediction of the coding sequences of unidentified human genes. XII.
RA The complete sequences of 100 new cDNA clones from brain which code
RA for large proteins in vitro.";
RL DNA Res. 5:355-364(1998).
RN [2]
RP SEQUENCE OF 1-253 FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
RA Tani H., Kinata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K.,
RA Masuho Y., Kanehori K.;
RA "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]

```

63  AATCCCGCCAGACGGAGAGCGTCTTATAATGTTTGCACAACTGTGTAT 112
    :::::||||| ||| ||||| ||||| :::::
272  AspProAlaValArgGlyLeuThrGluCysLeuGlnThrIle..... 286
    :::::||||| :::::
113  TCTATATGTTGAATCTGTACAGCAAAACCAACACAGAAGCTCCAAA 162
    ||||| :::::||||| :::::
287  .....LeuAsnLysAlaGlnGluProProLysSerLysLysValGlnH 301
    :::::
163  CAAGCAAAATCAAAAT 179
    ||| :::: |||||
301  iSerAsnAlaLysAsn 306

```

This Page Blank (uspto)

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-792, 'ES' <EL3>
A:Cross-references: EMBL:AF001262; NID:q2623764; PIDN:AAB86530.1; PID:q2623765
A:Experimental source: lung

alignment_scores:
Quality: 160.00 Length: 58
Ratio: 3.902 Gaps: 0
Percent Similarity: 70.690 Percent Identity: 51.724

alignment_block:

US-09-049-696-4 x T02205 ..

Align seg 1/1 to: T02205 from: 1 to: 905

8 TGCACATTCACAACTACAGGACTCTATGAAAGAGTGTGAGTTTGT 57

211 CysArgArgAspSerGlnThrGlyLeuTyrGluAlaLysCysThrPheLe 227

58 TCTCCAAATCCGCCAGACGAGAGGCTTCTATATGTTTGCACACATG 107

227 uProLysLysSerGlnThrAlaLysGluSerIleMetPheMetProSerL 244

108 TTGATTCTTATAGTTGAATTCCTGTACAGACAAACACCAACAAAGAGCT 157

244 euHisSerValThrGluPheCysThrGluLysThrHisAsnThrGluAla 260

158 CCNAACAGCAAAATCAAAATGCC 181

261 ProAsnLeuGlnAsnLysMetCys 268

seq_name: pir2:T28156

seq_documentation_block:
DNA-directed RNA polymerase homolog - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28156
R:Fox, B.A.; Li, W.B.; Tanaka, M.; Inselburg, J.; Bzik, D.J.
Mol. Biochem. Parasitol. 61, 37-48, 1993
A:Title: Molecular characterization of the largest subunit of plasmodium falciparum RNA
A:Reference number: Z20478; MUID:94081864

A:Accession: T28156

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2910 <FOX>

A:Cross-references: EMBL:L11172; NID:q414321; PID:q414322; PIDN:AAA72349.1

C:Genetics:

A:Introns: 2840/3; 2880/1

alignment_scores:
Quality: 66.00 Length: 99
Ratio: 1.535 Gaps: 4
Percent Similarity: 43.434 Percent Identity: 24.242

alignment_block:

US-09-049-696-4 x T28156 ..

Align seg 1/1 to: T28156 from: 1 to: 2910

2 AAAGATGCACATTCATTAAGTAACAGGACTCTATGAAAGAGTGTGA 51

1292 LysLysLeuIleLeuAsnLysMetTyrAspLeuTyrGluLysGlnAsnG 1308

52 GTTTGTTT.....CTCC 62

1308 uLeuIleAsnAsnPheAsnGlnLysGluAsnTyrLeuAsnLysMetG 1325

63 AATCCCGCCAGACGAGAGGCTTCT...ATAATGTTTCACACATGTT 109

.....

1325 LuLysArgAsnIleLysLysAspSerHisValMetLysAspGluHisIle 1341

110 GATTCTATA..... 118

1342 LeuSerIleSerGluAsnSerTyrLeuLeuAlaLeuLeuAsnLysGluLe 1358

119GTTGAATTCGTACACAAC 137

1358 uGlnAsnValIleTyrHisLeLeuAsnGluValGlnPheCysAspAspA 1375

1138 AAAACCCAC.....AACAAAGAGAGCTCCAAACAGCAAAATCAAAA 178

1375 spAlaSerLeuLeuAsnLysGluLysAsnGluLysGluAsnGluLys 1390

seq_name: pir2:T04229

seq_documentation_block:

ABC-type transport protein F14M19.30 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 17-Mar-2000

C:Accession: T04229

R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Hohelsel, J.;

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15262

A:Accession: T04229

A:Molecule type: DNA

A:Residues: 1-577 <BEV>

A:Cross-references: EMBL:AL049480

A:Experimental source: cultivar Columbia; BAC clone F14M19

C:Genetics:

A:Map position: 4

A:Note: F14M19.30

C:Superfamily: fruit fly white protein; ATP-binding cassette homology

alignment_scores:

Quality: 63.50 Length: 58

Ratio: 1.671 Gaps: 3

Percent Similarity: 65.517 Percent Identity: 29.310

alignment_block:

US-09-049-696-4/rev x T04229 ..

Align seg 1/1 to: T04229 from: 1 to: 577

181 GCATTTTTCATTTTCGCTTGTGGAGCTCTTGTGTGGTTTGTCTG 132

430 AlaTyrPheValLeuValIleTyr.....IleIleValLeuMetal 443

131 TACAGAAATTCACATATACAAATCAACATCTGTGCAACATTTATAGAGCC 82

443 aAsnSerPheValLeuPheLeuSerSerLeuAlaProAsnTyr..... 457

81 TTCTCCGTCGCGGGATTGGAGAAC.....AACTCACATCCTTT 41

458IleAlaGlyThrSerSerValThrIleLeuLeuAlaAlaPhe 471

40 TTCATAGAGTCCTGTTACTTTATT 17

472 PheLeuPheSerGlyTyrPheIle 479

seq_name: pir2:G86201

seq_documentation_block:

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: G86201

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansan, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

93 ATTATAGAAGCTTCCTCGTGTGGGATTGGAGAACAAACATCC 44
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 214 ISSerArgSerLeuLeuAsnLeuTyrGlyValAspLysLysSerProSer 230
 43 TTTTTCATA 35
 ||| |||
 231 PheArgile 233

seq_name: pir2:A58947

seq_documentation_block:

signal recognition particle 68K protein - dog

C:Species: Canis lupus familiaris (dog)

C>Date: 16-Apr-1999 #sequence_revision 16-Apr-1999 #text_change 08-Oct-1999

C:Accession: A58947; SI2981

R:Luetteke, H.

submitted to the EMBL Data Library, October 1992

A:Reference number: A58947

A:Accession: A58947

A:Molecule type: mRNA

A:Residues: 1-616 <LUE>

A:CROSS-references: GB:X53744; NID:g9331; PIDN:CAA37773.1; PID:g9332

A:Experimental source: kidney MDCK cells

R:Herz, J.; Flint, N.; Stanley, K.; Frank, R.; Dobberstein, B.

FEBS Lett. 276, 103-107, 1990

A>Title: The 68 kDa protein of signal recognition particle contains a glycine-rich region

A:Reference number: SI2981; MUID:91092392

A:Accession: SI2981

A:Molecule type: mRNA

A:Residues: 1-17, 'GGGGGG', 18-616 <HER>

A:CROSS-references: GB:X53744; NID:g9331

A>Note: this sequence corrected by submission in reference A58947

alignment_scores:

Quality: 62.00 Length: 57
 Ratio: 1.771 Gaps: 2
 Percent Similarity: 61.404 Percent Identity: 29.825

alignment_block:

US-09-049-696-4 x A58947 ..

Align seg 1/1 to: A58947 from: 1 to: 616

14 TTCATAAAGTAACAGACTCTATGAAGAAGGATGTCAGTTTGTCTCCA 63
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 202 PheAsnLysCysLysThrIleTyrGluLys.....LeuAl 213
 64 ATCCCGCCGACGAGGAGGCTTCTATAATGTTTGCAACATGTTGATT 113
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 213 aSerAlaPheThrGluGluGlnAlaValLeuTyrAsnGlnArgValGluG 230
 114 CTATA.....GTTGAATTCGTACAGAACAAACCAACAAAGAA 154
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 230 lulleSerProAsnIleArgTyrCysAlaTyrAsnIleGlyAspGlnSer 246
 155 GCTCCAAACAGCAATCAA 175
 ||| ||||| |||||
 247 AlaIleAsnGluLeuMetGln 253

seq_name: pir1:JDVLHH

seq_documentation_block:

DNA-directed DNA polymerase (EC 2.7.7.7) - heron hepatitis virus

C:Species: heron hepatitis virus, HHBV

A>Note: host Ardea cinerea (gray heron)

C>Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 11-Jun-1999

C:Accession: A30082

R:Sprengel, R.; Kaleta, E.F.; Will, H.

J. Virol. 62, 3832-3839, 1988

A>Title: Isolation and characterization of a hepatitis B virus endemic in herons.

A:Reference number: A93037; MUID:88333160

A:Accession: A30082

A:Molecule type: DNA
 A:Residues: 1-788 <SPR>
 A:CROSS-references: GB:M2056; NID:g325452; PIDN:AAA45738.1; PID:g325454
 C:Superfamily: hepatitis virus DNA-directed DNA polymerase
 C:Keywords: DNA biosynthesis; nucleotidyltransferase

alignment_scores:

Quality: 61.00 Length: 55
 Ratio: 1.743 Gaps: 2
 Percent Similarity: 63.636 Percent Identity: 32.727

alignment_block:

US-09-049-696-4 x JDVLHH ..

Align seg 1/1 to: JDVLHH from: 1 to: 788

5 AGATGCACATTCATATAAGTAACAGGACTCTATGAA...AAAGGATGTA 51
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 86 ArgHisLysLeuGlyLysLeuThrGlyLeuTyrGlnMetLysGlyCysG 102
 52 GTTGTCTCCAAATCCCGCCAGAGGAGGCTTCTATAATGTTGCAC 101
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 102 uPheAsnProHisTrpLysIleProAspIleSerAlaThrAsnPheSerG 119
 102 ACATGTTGATTCATAGTTGAATTCGTACAGAACAAACCAACAAA 151
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 119 In.....GluIleIleAsnGluCysProSerArgAsnTrpLysTyr 132
 152 GAAGCTCCAAACAAAG 166
 ||||| |||||
 133 LeuThrProAlaLys 137

seq_name: pir1:GNNYB4

seq_documentation_block:

genome polyprotein - coxsackievirus B4

N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core

polymerase (EC 2.7.7.48)

C:Species: coxsackievirus B4

C>Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 16-Jul-1999

C:Accession: A27170; A45576; B45576

R:Jenkins, O.; Booth, J.D.; Minor, P.D.; Almond, J.W.

J. Gen. Virol. 68, 1835-1848, 1987

A>Title: The complete nucleotide sequence of coxsackievirus B4 and its comparison to

A:Reference number: A27170; MUID:87253111

A:Accession: A27170

A:Molecule type: genomic RNA

A:Residues: 1-2183 <JEN>

R:Ramisigh, A.; Araki, H.; Bryant, S.; Hixson, A.

Virus Res. 23, 281-292, 1992

A>Title: Identification of candidate sequences that determine virulence in Cocksackiev

A:Reference number: A45576; MUID:92327833

A:Accession: A45576

A:Molecule type: genomic RNA

A:Residues: 569-696, 'T', 698-852 <RAM>

A:CROSS-references: GB:S39291; NID:g250908; PIDN:AAB22445.1; PID:g250909

A:Experimental source: B4, virulent strain

A>Note: sequence inconsistent with the nucleotide translation

A>Note: sequence extracted from NCBI backbone (NCBIN:108105, NCBIP:108106)

A:Accession: B45576

A:Molecule type: genomic RNA

A:Residues: 70-203, 'A', 205-231, 'E', 233-269, 'A', 271-330 <RA2>

A:CROSS-references: GB:S39291; NID:g250908; PIDN:AAB22446.1; PID:g250910

A:Experimental source: B4, virulent strain

A>Note: sequence extracted from NCBI backbone (NCBIN:108105, NCBIP:108107)

C:Superfamily: poliovirus genome polyprotein

C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransferase

F:1-69/Product: coat protein 1A #status predicted <CIA>

F:70-330/Product: coat protein 1B #status predicted <CIB>

F:331-568/Product: coat protein 1C #status predicted <CIC>

F:569-852/Product: coat protein 1D #status predicted <CID>

F:853-999/Product: core protein 2A #status predicted <C2A>

OM of: US-09-049-696-4 to: Issued_Patents_AA:* out_format : pfs

Date: Mar 30, 2002 2:32 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=framet_n2p.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09049696/runat_28032002_145238_2012/app_query.fasta_1.12579
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -GAPEXT=0.000
-EGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-EGAPOP=6.000 -EGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696@cgn1_1109 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT THREADS=1
```

Search information block:

```
Query: US-09-049-696-4
Query length: 181
Database: Issued Patents_AA:*
Database sequences: 212252
Database length: 22503292
Search time (sec): 362.110000
```

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-477-451-16 -			66.00	129.16	0.5408
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-219-237B-8 +			58.00	134.10	3.69
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-477-347-17 +			58.00	134.03	3.70
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-476-862-8 +			58.00	134.03	3.70
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-974-022-51 +			58.00	130.25	4.10
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-795-445A-51 +			58.00	130.25	4.10
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-795-447A-51 +			58.00	130.25	4.10
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-974-186-51 +			58.00	130.25	4.10
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-795-446B-51 +			56.50	123.11	7.31
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-915-934-2 -			56.50	123.11	7.31
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-325-743-2 -			56.50	109.46	10.60
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-220-081-2 +			56.50	102.39	12.85
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-755-587-16 +			56.50	98.59	14.25
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-639-501-2 +			56.50	98.59	14.25
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-603-753D-4 +			56.50	98.59	14.25
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-09-044-946-2 +			56.50	98.59	14.25
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-755-587-44 +			56.50	98.59	14.25
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-09-044-908-2 +			56.50	98.59	14.25
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-099-753-4 +			56.50	98.59	14.25
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-986-106-4 +			56.50	98.59	14.25
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-116-498-10 -			56.00	119.60	9.15
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-045-583-48 -			56.00	117.22	9.24
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-437-046B-67 +			56.00	117.39	9.72
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-685-259-25 -			55.50	103.06	16.31
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-762-500-25 -			55.50	103.06	16.31
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-762-500-75 -			55.50	102.94	16.37
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-172-422-1 -			55.00	97.69	21.47
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-172-422-1 -			55.00	95.43	22.83
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-477-451-16 -			54.50	114.72	15.36
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-517-802-3 +			54.00	109.74	19.99
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-272-414-2 +			54.00	107.18	21.43
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-07-861-458C-99 +			54.00	105.51	22.43
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-117-562-4 +			54.00	105.51	22.43
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-486-924-1 +			53.00	113.11	23.58
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-447-185-1 +			53.00	111.44	24.68
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-172-841-53 +			52.50	118.98	22.86
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-248-466B-9 +			52.50	105.66	32.84
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-910-551B-2 +			52.50	104.72	33.65
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-781-891-77 +			52.50	98.18	40.25
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-937-236-3 +			52.50	98.17	40.26
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-937-236-6 +			52.50	98.17	40.26

```
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-569-214-3 + 52.50 98.06 40.38
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-937-236-2 + 52.50 98.06 40.38
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-569-214-5 + 52.50 98.06 40.39
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-569-214-6 + 52.50 98.06 40.39
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-937-236-5 + 52.50 98.06 40.39
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-477-451-16
```

seq_documentation_block:

```
: Sequence 16, Application US/08477451
: Patent No. 5928865
: GENERAL INFORMATION:
: APPLICANT: Covacci, Antonello
: TITLE OF INVENTION: 46
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Chiron Corporation
: STREET: 4560 Horton Street
: CITY: Emeryville
: STATE: CA
: COUNTRY: USA
: ZIP: 94608-2916
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/477,451
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: McClung, Barbara G.
: REGISTRATION NUMBER: 33,113
: REFERENCE/DOCKET NUMBER: 0335.002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 510-601-2708
: TELEFAX: 510-655-3542
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1786 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-477-451-16
```

```
alignment_scores:
Quality: 66.00 Length: 77
Ratio: 1.886 Gaps: 4
Percent Similarity: 45.455 Percent Identity: 29.870
```

alignment_block:

```
US-09-049-696-4/rev x US-08-477-451-16 ..
Align seg 1/1 to: US-08-477-451-16 from: 1 to: 1786
178 TTTTGTGATTTTGGCTTTGGAGCTTCTTTCTT .....GTGGTTT 138
||||| : : : : : ||||| : : : : :
138 PheLeuSerPheValLeuSerArgPhePheLeuAsnArgValValLe 1401
137 GTTCTGTACAAATCAATATAGAAATCAATGTTGTGCAACATTATA 88
||||| : : : : : ||||| : : : : :
1401 uPhePheLysArgPheAsnLeuLeuAsnLeuLeuCys ..... 1414
87 GAAGCCTTCTCCGCTCGGGGATTGGAGAACAACTCACATCTCTTTC 38
||||| : : : : : ||||| : : : : :
1415 .....SerArgProIleIleSerPhePhe 1422
37 ATAGAGTCC .....TGTTACTTT ..... 20
1423 SerGlnSerValSerGlyGluPhePhePheLeuPheCysAsnPheVa 1439
```

```

seq_documentation_block:
; Sequence 17, Application US/08477347
; Patent No. 6232446
; GENERAL INFORMATION:
; APPLICANT: WALLACH, David
; APPLICANT: BIGDA, Jacek
; APPLICANT: BELETSKY, Igor
; APPLICANT: METT, Igor
; TITLE OF INVENTION: TNF LIGANDS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,347
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/115,685
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 106271
; FILING DATE: 08-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, G. Kevin
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: WALLACH=10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;S-08-477-347-17

```

alignment_scores:			
Quality:	58.00	Length:	41
Ratio:	2.417	Gaps:	3
Percent Similarity:	58.537	Percent Identity:	41.463

```
alignment_block:
US-09-049-696-4 x US-08-477-347-17 ..
```

Align seq 1/1 to: US-08-477-347-17 from: 1 to: 140

52 GTTTGTTCTCCATCCCGCCACAGCGGACTTCTATAATGTTTGAC 101
||||| ||||
34 ValCysHisProCysGluPro.....GlyPheTrpAsnGlu..Alav 47

102 AACATTGTGATTCTATATAGTTGAATTCGTACAGACAACCAACACAAA 151
::: |||||:
47 alaAsnIyrAspThrCysLysGln...CystHrGlnCysAsnHisArgSer 62

152 GAAGCTCCAAACAAGCAAAAT 172
 :: |||||
 63 GlySerGluLeuLysGlnAsn 69

seq_name: /cqn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-476-862-8

[illegible]

19


```
;
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,186
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/577,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-974-186-51

alignment_scores:
    Quality: 58.00      Length: 41
    Ratio: 2.417        Gaps: 3
    Percent Similarity: 58.537      Percent Identity: 41.463

alignment_block:
US-09-049-696-4 x US-08-974-186-51 ..
Align seg 1/1 to: US-08-974-186-51 from: 1 to: 205

52 GTTGTGTTCTCAATCCGCGCAGAGGAGGCTTCTATAATGTTGCAC 101
   ||||| ||| ||| ||||| ||||| |||
58 ValCysHisProCysGluPro.....GlyPheTyrAsnGlu.Alav 71

102 AACATGTTGATTCATATAGTTGTAATCTGTACAGAACAAACCAACAAA 151
   :: ||||| :: ||||| :: ||||| :: ||||| :: ||||| ::
71 alAsnTyrAspThrCysLysGln...CysThrGlnCysAsnHisArgSer 86

152 GAAGCTCCAAACAAACGCAAAAT 172
   ::||| ||||| |||||
87 GlySerGluLeuLysGlnAsn 93

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pap:US-08-795-446B-51

seq_documentation_block:
; Sequence 51, Application US/08795446B
; Patent No. 6288032
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/795,446B
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/577,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-795-446B-51

alignment_scores:
    Quality: 58.00      Length: 41
    Ratio: 2.417        Gaps: 3
    Percent Similarity: 58.537      Percent Identity: 41.463

alignment_block:
US-09-049-696-4 x US-08-795-446B-51 ..
Align seg 1/1 to: US-08-795-446B-51 from: 1 to: 205

52 GTTGTGTTCTCAATCCGCGCAGAGGAGGCTTCTATAATGTTGCAC 101
   ||||| ||| ||| ||||| ||||| |||
58 ValCysHisProCysGluPro.....GlyPheTyrAsnGlu.Alav 71

102 AACATGTTGATTCATATAGTTGTAATCTGTACAGAACAAACCAACAAA 151
   :: ||||| :: ||||| :: ||||| :: ||||| :: ||||| ::
71 alAsnTyrAspThrCysLysGln...CysThrGlnCysAsnHisArgSer 86

152 GAAGCTCCAAACAAACGCAAAAT 172
   ::||| ||||| |||||
87 GlySerGluLeuLysGlnAsn 93

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pap:US-07-915-934-2

seq_documentation_block:
; Sequence 2, Application US/07915934
; Patent No. 5360893
; GENERAL INFORMATION:
; APPLICANT: OWENS, GREGORY P.
; APPLICANT: COHEN, J. J.
; APPLICANT: HAHN, WILLIAM F.
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PROTEINS USED TO
; TITLE OF INVENTION: ELICIT AND DETECT PROGRAMMED CELL DEATH
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FREDERICK W. PEPPER, PH.D.
; STREET: 11545 W. BERNARDO COURT, STE. 302
; CITY: SAN DIEGO
; STATE: CA
; COUNTRY: USA
; ZIP: 92127
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/915,934
; FILING DATE: 19920720
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PEPPER PH.D., FREDERICK W.
; REGISTRATION NUMBER: 31,286
; REFERENCE/DOCKET NUMBER: 92-0224.02
```

```

/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 287 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-325-743-2

alignment_scores:
  Quality: 56.50      Length: 38
  Ratio: 2.260       Gaps: 1
Percent Similarity: 65.789      Percent Identity: 36.842

alignment_block:
US-09-049-696-4/rev x US-08-325-743-2 ..

Align seg 1/1 to: US-08-325-743-2 from: 1 to: 287

179 ATTTTTCATTTGCTTGTGGAGCTTCTTTGTTGT...GGTTTGTGTTCT 133
    ::::| | | | | | | | | | | | | | | | | | | | | | | |
23 LeuPheLeuPheCysCysArgGluProLeuCysCysAlaGlyLeuArgVa 39

122 GTACAGAAATTCACCTATAGAAATCAACATGTTGTGCAACATATTAGAAGC 83
    ::::| | | | | | | | | | | | | | | | | | | | | | | |
39 lPheArgAsnGlnLeuProArgLysAsnAlaPheTyrSerTyrGluProP 56

82 CTTCCTCGTCTGGC 69
    | | | | |
56' roserGluThrGly 60

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-220-081-2

seq_documentation_block:
/ Sequence 2, Application US/092200081
/ Patent No. 6171833
/ GENERAL INFORMATION:
/ APPLICANT: Sinskey, Anthony J.
/ APPLICANT: Lessard, Phillip A.
/ APPLICANT: Willis, Laura B.
/ APPLICANT: Stephanopoulos, Gregory
/ TITLE OF INVENTION: Pyruvate Carboxylase from Corynebacterium glutamicum
/ FILE REFERENCE: 1533.0790000
/ CURRENT APPLICATION NUMBER: US/09/220.081
/ CURRENT FILING DATE: 1998-12-23
/ NUMBER OF SEQ ID NOS: 36
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 2
/ LENGTH: 1140
/ TYPE: PRT
/ ORGANISM: Corynebacterium glutamicum
US-09-220-081-2

alignment_scores:
  Quality: 56.50      Length: 46
  Ratio: 2.093       Gaps: 2
Percent Similarity: 58.696      Percent Identity: 34.783

alignment_block:
US-09-049-696-4 x US-09-220-081-2 ..

Align seg 1/1 to: US-09-220-081-2 from: 1 to: 1140

20 AAGTACAGGACTCATGAAAGAGATGTGATTTGTTCTCCAAATCCCG 69
    ::::| | | | | | | | | | | | | | | | | | | | | | | |
230 GluValValHisLeuTyrGluArgAspCysSer.....LeuGlnArgAr 244

79 CCAGACGGACGAGCTTCATATGTTTTCACAAATGTTGAT..... 112
    | | | | | : : : : : | | | | | | | | | | | | | | | |
244 gHisGlnLysValGluIleAlaProAlaGlnHisLeuAspProGluL 261

113 .....TCTATATGTAATTTCTGT 130
    | | | | | | | | | | | | | | | | | | | | | | | |

```



```
alignment_block:
US-09-049-696-4 x US-08-639-501-2
Align seg 1/1 to: US-08-639-501-2 from: 1 to: 3418
9 GCACATTCATTAAGTAACAG.....GACTCTATCAAAAGGATGT 49
1222 AlaHisGlyThrLysLeuAsnValSerThrGluAlaLeuGlnLysAlaVa 1238
50 GAGTTTGTCTCCAA.....TCCCGCCAGA 74
1238 LlysLeuPheSerAspIleGluAsnIleSerGluGluThrSerAlaGluV 1255
75 CGAGAAGCCTCTATAATGTTGGACACATGTTGATTCATAGTTGAA 124
1255 aHisProileSerLeuSerSerLysCysHisAspValValSer 1271
125 TTCTGTACAGAACAAACACACAAAGAGCTCCAAACGAAATCA 174
1272 MetPheLysIleGluAsnHisAsnAspLysThrValSerGluLysAsnAs 1288
175 AAAATGC 181
1288 nLysCys 1290
seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.US-08-603-753D-4
seq_documentation_block:
; Sequence 4, Application US/08603753D
; Patent No. 5891857
; GENERAL INFORMATION:
; APPLICANT: HOLT, JEFFREY T.
; APPLICANT: JENSEN, ROY A.
; APPLICANT: PAGE, DAVID L.
; APPLICANT: KING, MARY-CLAIRE
; APPLICANT: SZABO, CSILLA I.
; APPLICANT: JETTON, THOMAS L.
; APPLICANT: ROBINSON-BENION, CHERYL L.
; APPLICANT: THOMPSON, MARILYN E.
; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2
; TITLE OF INVENTION: PROTEINS AND SCREENING AND THERAPEUTIC METHODS BASED ON
; TITLE OF INVENTION: CHARACTERIZED BRCA1 AND BRCA2 PROTEINS.
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARLES A. TAYLOR, JR.
; STREET: SUITE 1401, UNIVERSITY TOWER, 3100 TOWER
; STREET: BOULEVARD
; CITY: DURHAM
; STATE: NORTH CAROLINA
; COUNTRY: USA
; ZIP: 27707
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage
; COMPUTER: IBM PC/XT/AT compatible
; OPERATING SYSTEM: Windows 3.1
; SOFTWARE: WORD PERFECT 6.1 and ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/603,753D
; FILING DATE: 20 FEB 1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. 08/373,799
; FILING DATE: 17 JAN 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ARLES A. TAYLOR, JR.
; REGISTRATION NUMBER: 39,395
; REFERENCE/DOCKET NUMBER: 1242/2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 493-8000
; TELEFAX: (919) 419-0383
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
LENGTH: 3418
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHEICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Homo sapiens sapiens
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE: adult
TISSUE TYPE: female breast
CELL TYPE: normal breast tissue
CELL LINE: HMEC
ORGANELLE: no
FEATURE:
NAME/KEY: BRCA2 protein
LOCATION: 1 to 3418; Genbank locus HSU43746
IDENTIFICATION METHOD:
OTHER INFORMATION: BRCA2 protein has a negative
PUBLICATION INFORMATION:
AUTHORS: Wooster, R. et al.
TITLE: Identification of the breast cancer
GENE: susceptibility gene BRCA2
JOURNAL: Nature
VOLUME: 379
PAGES: 789-792
DATE: 1995
RELEVANT RESIDUES IN SEQ ID NO: 4: granin box
RELEVANT RESIDUES IN SEQ ID NO: domain at amino acids 3334-3344
US-08-603-753D-4
alignment_scores:
Quality: 56.50 Length: 69
Ratio: 1.314 Gaps: 2
Percent Similarity: 62.319 Percent Identity: 27.536
alignment_block:
US-09-049-696-4 x US-08-603-753D-4
Align seg 1/1 to: US-08-603-753D-4 from: 1 to: 3418
9 GCACATTCATTAAGTAACAG.....GACTCTATCAAAAGGATGT 49
1222 AlaHisGlyThrLysLeuAsnValSerThrGluAlaLeuGlnLysAlaVa 1238
50 GAGTTTGTCTCCAA.....TCCCGCCAGA 74
1238 LlysLeuPheSerAspIleGluAsnIleSerGluGluThrSerAlaGluV 1255
75 CGAGAAGCCTCTATAATGTTGGACACATGTTGATTCATAGTTGAA 124
1255 aHisProileSerLeuSerSerLysCysHisAspValValSer 1271
125 TTCTGTACAGAACAAACACACAAAGAGCTCCAAACGAAATCA 174
1272 MetPheLysIleGluAsnHisAsnAspLysThrValSerGluLysAsnAs 1288
175 AAAATGC 181
1288 nLysCys 1290
```

OM of: US-09-049-696-4 to: A_Geneseq_1101.* out_format : pfs
Date: Mar 30, 2002 2:18 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL-frame_h2p_model -DEV-xlp
-Q=/cgn2.1/USPTO.spool/US09049696/runat_28032002_145237_2001/app_query.fasta_1.12579
-DB=A_Geneseq_1101 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 @CGN1_1_293 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-049-696-4
Query length: 181
Database: A_Geneseq_1101.*
Database sequences: 522463
Database length: 74073290
Search time (sec): 819.910000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA675614 +		324.00	785.61	869	1.0e-36
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AA674824 +		324.00	785.13	914	1.0e-36
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA624514 +		324.00	785.13	914	1.0e-36
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA673716 +		324.00	785.13	914	1.0e-36
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA674733 +		324.00	785.13	914	1.0e-36
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA675474 +		324.00	785.02	925	1.0e-36
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AA673715 +		210.00	497.96	913	1.0e-20
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA673715 +		210.00	497.96	913	1.0e-20
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA66749 +		180.00	422.32	919	1.6e-16
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA687560 +		180.00	422.32	919	1.6e-16
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA65272 +		180.00	422.32	919	1.6e-16
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AA6741039 +		156.50	367.28	592	3.0e-13
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA611324 +		156.50	367.28	592	3.0e-13
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA645904 +		156.50	365.14	742	3.1e-13
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AA641040 +		156.50	364.54	791	3.1e-13
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA611325 +		156.50	364.54	791	3.1e-13
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AA641036 +		156.50	362.88	943	3.3e-13
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA611321 +		156.50	362.88	943	3.3e-13
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AA674823 +		152.50	352.80	943	1.2e-12
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA670459 +		121.00	273.46	942	3.1e-08
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA624943 +		63.50	133.95	535	3.24
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA624942 +		63.50	133.27	575	3.29
/SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT:AA612141 +		63.00	119.39	2185	5.14
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA693530 +		62.50	152.44	369	2.79
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA693530 +		62.00	133.69	438	4.87
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA64394 +		62.00	132.07	505	5.05
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA64394 +		62.00	128.80	543	5.43
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA641697 +		62.00	128.78	620	5.43
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA639911 +		62.00	128.67	627	5.45
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA642724 +		61.00	133.19	298	6.42
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA666357 +		61.00	133.09	301	6.44
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA671414 +		61.00	132.78	311	6.48
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA672361 +		61.00	132.78	311	6.48
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA694034 +		61.00	132.45	928	8.18
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA67451 +		61.00	122.45	928	8.18
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA609067 +		60.00	131.39	276	8.77
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA609066 +		60.00	131.19	282	8.77
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA609065 +		60.00	129.65	332	9.08
/SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AA664193 +		59.50	112.32	1817	15.32
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA693518 +		59.00	126.55	353	12.71

/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA688483 + 59.00 123.90 13.49 467
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA694819 + 59.00 120.69 14.50 656
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA642227 + 59.00 117.40 15.61 929
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA639131 + 59.00 117.30 15.65 939
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA640917 + 59.00 116.91 15.78 978

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA675614

seq_documentation_block:

ID AA675614 standard; Protein; 869 AA.
XX
AC AA675614;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6378.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX

PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI: 2001-235357/24.
DR N-PSDB; AAH35019.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 11; Page 7851-7854; 9803pp; English.

AAH32943 to AAH37195 and AA673514 to AA677788 represent human colon
cancer-associated nucleic acid molecules (N) and proteins (P), where
the proteins are collectively known as colon cancer antigens. The colon
cancer antigens have cytostatic activity and can be used in gene
therapy and vaccine production. N and P may be used in the prevention,
diagnosis and treatment of diseases associated with inappropriate P
expression. For example, N and P may be used to treat disorders
associated with decreased expression by rectifying mutations or deletions
in a patient's genome that affect the activity of P by expressing
inactive proteins or to supplement the patients own production of P.
Additionally, N may be used to produce the colon cancer-associated Ps,
by inserting the nucleic acids into a host cell and culturing the cell
to express the proteins. N and P can be used in the prevention, diagnosis
and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
and AA677789 represent sequences used in the exemplification of the
present invention.
N.B. Pages 666 to 682 and page 7053 of the sequence listing were
missing at time of publication, meaning no sequences are present for
SEQ ID NO:1027 to 1052, 7921 and 7922.

Sequence 869 AA;

alignment_scores:
Quality: 324.00 Length: 60
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
*US-09-049-696-4 x AA675614

Align seg 1/1 to: AAG75614 from: 1 to: 869

```

2 AAAGATGCACATTCAATAAGTAACAGAGCTCTATGAAAAGGATGTGA 51
|||||
163 LysArgCysThrPheAsnLysValThrGlyLeuTyrGluLysGlyCysG1 179
52 GTTGTCTCCCAATCCCGCCAGAGAGAGGCTTCTATATGTTTGCAC 101
|||||
179 uPheValLeuGlnSerArgGlnThrGluLysAlaSerIleMetPheAlaG 196
|||||
102 AACATGTTGATTCTATAGTTGAATCTGTACAGAAACAAAACACACAAA 151
|||||
196 lnHisValaspSerIleValGluPheCysThrGluGlnAsnHisAsnLys 212
152 GAAGTCTCCAAACAGCAAAATCAAAAATGTC 181
|||||
213 GluAlaProAsnLysGlnAsnGlnLysCys 222

```

seq_name: /SIDS2/gcgdata/geneseq/AA1999.DAT: AAB74824

seq_documentation_block:

ID AAB74824 standard; Protein: 914 AA.

XX AC AAB74824;

DT 13-JUN-2001 (first entry)

XX DE Human ICACC-1 protein sequence.

XX KW ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
 KW interleukin 9 induced calcium activated chloride channel; IL-9;
 KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
 KW anti-inflammatory; immunomodulatory; cystic fibrosis;
 KW inflammatory bowel disease; autoimmune disease.

XX OS Homo sapiens.

XX PN WO9944620-A1.

XX PD 10-SEP-1999.

XX PF 03-MAR-1999; 99WO-US04703.

XX PR 03-MAR-1998; 98US-0076815.

XX PA (MAGA-) MAGAININ PHARM INC.

XX PI Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;

XX PI Nicolaides NC, Zhou Y, Dong Q;

XX DR WPI; 1999-550979/46.

XX DR N-PSDB; AAF81927.

XX PT New nucleic acid encoding calcium activated chloride channel, used to
 PT identify, e.g. specific modulators for treating atopic allergy -

XX PS Claim 11; Fig 4B; 75pp; English.

XX CC The present sequence represents the human interleukin 9 (IL-9) induced
 CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
 CC have anti-allergic, anti-asthmatic, anti-inflammatory and
 CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
 CC used to alleviate asthma (or more generally atopic allergy), while those
 CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
 CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
 CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
 CC (increased levels) or IBD (reduced levels), also for monitoring
 CC treatment of these conditions. The ICACC proteins can be used:
 CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
 CC reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
 CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
 CC identify modulators and binding partners. ICACC polynucleotides can be

CC used to generate transgenic animals or recombinant cells, used to screen
 CC for antagonists, also as a source of therapeutic antisense agents or
 CC diagnostic probes (for quantifying mRNA expression, e.g. for
 CC identification of modulators).

XX SQ Sequence 914 AA;

alignment_scores:

Quality: 324.00 Length: 60
 Ratio: 5.400 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-4 x AAB74824 ..

Align seg 1/1 to: AAB74824 from: 1 to: 914

```

2 AAAGATGCACATTCAATAAGTAACAGAGCTCTATGAAAAGGATGTGA 51
|||||
208 LysArgCysThrPheAsnLysValThrGlyLeuTyrGluLysGlyCysG1 224
52 GTTGTCTCCCAATCCCGCCAGAGAGGCTTCTATATGTTTGCAC 101
|||||
224 uPheValLeuGlnSerArgGlnThrGluLysAlaSerIleMetPheAlaG 241
102 AACATGTTGATTCTATAGTTGAATCTGTACAGAAACAAAACACACAAA 151
|||||
241 lnHisValaspSerIleValGluPheCysThrGluGlnAsnHisAsnLys 257
152 GAAGTCTCCAAACAGCAAAATCAAAAATGTC 181
|||||
258 GluAlaProAsnLysGlnAsnGlnLysCys 267

```

seq_name: /SIDS2/gcgdata/geneseq/AA2001.DAT: AAM24514

seq_documentation_block:

ID AAM24514 standard; Protein: 914 AA.

XX AC AAM24514;

XX DT 12-OCT-2001 (first entry)

XX DE C902P predicted amino acid sequence.

XX KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
 KW immunogenic; gene therapy; vaccine; colonic cancer.

XX OS Homo sapiens.

XX PN WO200149716-A2.

XX PD 12-JUL-2001.

XX PF 29-DEC-2000; 2000WO-US355596.

XX PR 30-DEC-1999; 99US-0476296.

XX PR 10-JAN-2000; 2000US-0480321.

XX PR 15-FEB-2000; 2000US-0504629.

XX PR 06-MAR-2000; 2000US-0519444.

XX PR 19-MAY-2000; 2000US-0575251.

XX PR 29-JUN-2000; 2000US-0609448.

XX PR 28-AUG-2000; 2000US-0649811.

XX PA (CORI-) CORIXA CORP.

XX PI Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;

XX PI King GE, Wang T, Jiang Y;

XX DR WPI; 2001-441847/47.

XX CC Colon tumor associated proteins and nucleic acids useful for the
 PT prevention, diagnosis and treatment of colonic cancer -

```

XX PS Claim 2; Page 440-443; 472pp; English.
XX CC The present invention describes colon tumour associated proteins (I) and
XX CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX CC (II) may be used in the prevention, diagnosis and treatment of diseases
XX CC associated with inappropriate colon tumour associated protein (TCAP)
XX CC expression, such as colonic cancer. For example, (I) and (II) may be
XX CC used to treat disorders associated with decreased expression by
XX CC rectifying TCAPs by expressing inactive proteins or to supplement the
XX CC activity of TCAPs by deleting in a patient's genome that affect the
XX CC patients own production of them. Additionally, (II) may be used to
XX CC produce the TCAP proteins, by inserting the nucleic acids into a host
XX CC cell culturing the cell to express the protein. (II) and its
XX CC complementary sequences may also be used as DNA probes in diagnostic
XX CC polymerase chain reaction (PCR) and hybridisation assays to detect and
XX CC quantitate the presence of similar nucleic acids in samples, and
XX CC therefore which patients may be in need of restorative therapy. (I) may
XX CC also be used as antigens in the production of antibodies against TCAPs
XX CC and in assays to identify modulators of TCAP expression and activity.
XX CC Anti-(I) antibodies and antagonists may also be used to down regulate
XX CC TCAP expression and activity. The anti-(I) antibodies may also be used
XX CC as diagnostic agents for detecting the presence of TCAPs in samples
XX CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
XX CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
XX CC given in the exemplification of the present invention.
XX SQ Sequence 914 AA;

alignment_scores:
    Quality: 324.00 Length: 60
    Ratio: 5.400 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
    US-09-049-696-4 x AAM24514 ..

Align seg 1/1 to: AAM24514 from: 1 to: 914
    2 AAAAGATGCACATTCATTAAGTAACAGGACTCTATCAAAAAGCATGTGA 51
    |||||||
    208 LysArgCysThrPheAsnLysValThrGlyLeuTyrgLysGlyCysGI 224
    |||||||
    52 GTTGTGTCCTCAATCCCGCAGACGAGAGGCTTCTATTAATGTTTGCAC 101
    |||||||
    224 uPheValLeuGlnSerArgGlnThrGluLysAlaSerIleMetPheAlaG 241
    |||||||
    102 AACATGTTGATTCATAGTTGAATTCGTACAGAACAAACCAACACAAA 151
    |||||||
    241 InHisValAspSerIleValGluPheCysThrGluGlnAsnHisAsnLys 257
    |||||||
    152 GAAGCTCCAAACAAAGCAAAATCAAAATGC 181
    |||||||
    258 GluAlaProAsnLysGlnAsnGlnLysCys 267
    |||||||

seq_name: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT: AAB73716

seq_documentation_block:
    ID AAB73716 standard; Protein; 914 AA.
XX AC AAB73716;
XX AC AAB73716;
XX DT 11-SEP-2001 (first entry)
XX DE Human CLC1 protein, SEQ ID NO:2.
XX KW Human CLC1; goblet cell; mouse Gob-5 orthologue; drug screening;
XX KW expression inhibition; antisense therapy; gene therapy;
XX KW chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic.
XX OS Homo sapiens.

```

```

XX PH Key Location/Qualifiers
XX FT Misc-difference 152 /note= "Encoded by AGG in AAH46124"
XX EN WO200138530-A1.
XX RD 31-MAY-2001.
XX PF 22-NOV-2000; 2000WO-JP08232.
XX PR 24-NOV-1999; 99JP-0333479.
XX PR 27-APR-2000; 2000JP-0127589.
XX XX (TAKE ) TAKEDA CHEM IND LTD.
XX XX Nakanishi A, Morita S;
XX XX WPI; 2001-355935/37.
XX XX NPSDB; AAH46102, AAH46124.
XX XX New antisense nucleotide, useful for treatment and prevention of
XX XX bronchial asthma and chronic obstructive pulmonary disease -
XX FS Claim 2; Page 76-80; 104pp; Japanese.
XX CC This invention relates to an antisense nucleotide targetted to the mouse
XX CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
XX CC the CLC1 gene (coding sequence shown in AAH46102). The invention also
XX CC relates to an antibody specific for the Gob-5 protein, medical and
XX CC diagnostic compositions containing the antisense nucleotide or the
XX CC antibody, and methods and kits for screening for compounds which inhibit
XX CC the protein. Gob-5 and CLC1 are proteins expressed by goblet cells.
XX CC The antisense oligonucleotides and antibody are therefore useful for the
XX CC treatment and prevention of bronchial asthma and chronic obstructive
XX CC pulmonary disease. The present sequence represents human CLC1 protein.
XX SQ Sequence 914 AA;

alignment_scores:
    Quality: 324.00 Length: 60
    Ratio: 5.400 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
    US-09-049-696-4 x AAB73716 ..

Align seg 1/1 to: AAB73716 from: 1 to: 914
    2 AAAAGATGCACATTCATTAAGTAACAGGACTCTATCAAAAAGCATGTGA 51
    |||||||
    208 LysArgCysThrPheAsnLysValThrGlyLeuTyrgLysGlyCysGI 224
    |||||||
    52 GTTGTGTCCTCAATCCCGCAGACGAGAGGCTTCTATTAATGTTTGCAC 101
    |||||||
    224 uPheValLeuGlnSerArgGlnThrGluLysAlaSerIleMetPheAlaG 241
    |||||||
    102 AACATGTTGATTCATAGTTGAATTCGTACAGAACAAACCAACACAAA 151
    |||||||
    241 InHisValAspSerIleValGluPheCysThrGluGlnAsnHisAsnLys 257
    |||||||
    152 GAAGCTCCAAACAAAGCAAAATCAAAATGC 181
    |||||||
    258 GluAlaProAsnLysGlnAsnGlnLysCys 267
    |||||||

seq_name: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT: AAB74733

seq_documentation_block:
    ID AAB74733 standard; Protein; 914 AA.
XX AC AAB74733;
XX AC AAB74733;

```


alignment_scores:
Quality: 324.00 Length: 60
Ratio: 5.400 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-049-696-4 x AAG75474 ..
Align seg 1/1 to: AAG75474 from: 1 to: 925
2 AAGATGACATTCATTAAGTAAAGTAAAGGAGTGTGA 51
|||||
219 LysArgCysThrPheAsnLysValThrGlyLeuTyrgLulysGlyCysGI 235
52 GTTGTCTCCCAATCCCGCAGACGAGAGGCTTCTATATATGTTTGCAC 101
|||||
235 upheValLeuGlnSerArgGlnThrGluLysAlaSerIleMetPheAlaG 252
102 ACATGTTGATCTATAGTTGAATCTGTACAGACAAACACACACAAA 151
|||||
252 InHisValAspSerIleValGluPheCysThrGluGlnAsnHisAsnLys 268
152 GAAGCTCCAAACAAAGCAAAATCAAAATGC 181
|||||
269 GluAlaProAsnLysGlnAsnGlnLysCys 278

seq_name: /SIDS2/gcgdata/geneseq/AA1999.DAT: AAB74822

seq_documentation_block:

ID: AAB74822 standard; Protein: 913 AA.

AC AAB74822;

DT 13-JUN-2001 (first entry)

DE Murine ICACC-1 protein sequence.

XX ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW inflammatory bowel disease; autoimmune disease.
XX

OS Mus sp.

XX W09944620-A1.

XX 10-SEP-1999.

XX 03-MAR-1999; 99WO-US04703.

XX 03-MAR-1998; 98US-0076815.

XX (MAGA-) MAGAININ PHARM INC.

XX Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;

XX Nicolaides NC, Zhou Y, Dong Q;

XX WPI; 1999-550979/46.

XX N-PSDB; AAF81925.

XX New nucleic acid encoding calcium activated chloride channel, used to
PT identify, e.g. specific modulators for treating atopic allergy -
PS Claim 12; Fig 2; 75pp; English.

XX The present sequence represents the murine interleukin 9 (IL-9) induced
CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
CC have anti-allergic, anti-asthmatic, anti-inflammatory and
CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
CC used to alleviate asthma (or more generally atopic allergy), while those

CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
CC (increased levels) or IBD (reduced levels), also for monitoring
CC treatment of these conditions. The ICACC proteins can be used:
CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
CC reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
CC identify modulators and binding partners. ICACC polynucleotides can be
CC used to generate transgenic animals or recombinant cells, used to screen
CC for antagonists, also as a source of therapeutic antisense agents or
CC diagnostic probes (for quantifying mRNA expression, e.g. for
CC identification of modulators).

XX Sequence 913 AA;

alignment_scores:

Quality: 210.00 Length: 59

Ratio: 3.962 Gaps: 0

Percent Similarity: 89.831 Percent Identity: 61.017

alignment_block:

US-09-049-696-4 x AAB74822 ..

Align seg 1/1 to: AAB74822 from: 1 to: 913

5 AGATGCACATTCATTAAGTAAAGTAAAGGAGTGTGAGTT 54
|||||
219 LysCysValIleAspArgValThrGlyLeuTyrgLysAspAsnLysValPh 226
55 TGTCTCCCAATCCCGCAGACGAGAGGCTTCTATATGTTTGCACAC 104
|||||
225 eValProAspProHisGlnAsnGluLysAlaSerIleMetPheAsnGlnA 243
105 ATGTTGATCTATAGTTGAATCTGTACAGACAAACACACACAAAGAA 154
|||||
243 snIleAsnSerValValGluPheCysThrGluLysAsnHisAsnGlnGlu 259

155 GCTCCAAACAAAGCAAAATCAAAATGC 181

|||||

260 AlaProAsnAspGlnAsnGlnArgCys 268

seq_name: /SIDS2/gcgdata/geneseq/AA2001.DAT: AAB73715

seq_documentation_block:

ID: AAB73715 standard; Protein: 913 AA.

AC AAB73715;

DT 11-SEP-2001 (first entry)

DE Mouse Gob-5 protein, SEQ ID NO:1.

XX Mouse Gob-5; murine; goblet cell; human CLCA1 orthologue; drug screening;

XX expression inhibition; antisense therapy; gene therapy; bronchial asthma;
XX chronic obstructive pulmonary disease; antiasthmatic;
XX Mus sp.
XX W0200138530-A1.

XX 31-MAY-2001.

XX 22-NOV-2000; 2000WO-JP08232.

XX 24-NOV-1999; 99JP-0333479.

XX 27-APR-2000; 2000JP-0127589.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Nakanishi A, Morita S;

CC The invention provides membrane-bound PRO polypeptides and
 CC polynucleotides encoding them. The PRO sequences of the invention were
 CC identified based on extracellular domain homology screening. The PRO
 CC sequences have homology with proteins including LDL receptors, RTK,
 CC ligands and various enzymes. The membrane-bound proteins and receptor
 CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
 CC immunoadhesins, for instance, can be used as therapeutic agents to block
 CC receptor-ligand interactions. The membrane-bound proteins can also be
 CC employed for screening of potential peptide or small molecule inhibitors
 CC of the relevant receptor/ligand interaction. The PRO encoding sequences
 CC are useful as hybridization probes, in chromosome and gene mapping and in
 CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
 CC will also be useful for the preparation of PRO polypeptides, especially
 CC by recombinant techniques.
 XX

Seq Sequence 919 AA;
 Alignment_scores:
 Quality: 180.00 Length: 60
 Ratio: 3.750 Gaps: 0
 Percent Similarity: 80.000 Percent Identity: 55.000

alignment_block:
 .US-09-949-696-4 x AAY66749 ..
 Align seg 1/1 to: AAY66749 from: 1 to: 919

```

      2 AAAAGATGCACATTCAATAAAGTAACAGCAGCTCTATGAAAAGGATGTGA 51
      ||| |:| |:::|:::|::| |::| |::| |::|
    208 ArgAlaCysArgIleAspSerThrLysLeuTyrGlyLysAspCysG1 224
      52 GTTTCTTCCTCCAATCCCGCACGAGAGGCTTCTATATGTTTGAC 101
      ||| |:| |:::|:::|::| |::| |::| |::|
    224 nPhePheProaspLysValGlnThrGlulysAlaSerileMetPheMetG 241

    102 AACATGTTGATCTATAGTTGAATTCCTGACAGAACAAACCACAACAAA 151
      || |:| |:::|:::|::| |::| |::| |::|
    241 InSerIleAspSerValValGluPheCysAsnGluLysThrHisAsnGln 257

    152 GAAGCTCCAAACAGCAAATAACAAAAATGCC 181
      ||| |:| |:::|:::|::| |::| |::| |::|
    258 GluAlaProserLeuGlnAsnIleLysCys 267

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAB87560

seq_documentation_block:
ID AA87560 standard; Protein: 919 AA.
XX
AC AAP87560;
XX
XX
DT 15-MAY-2001 (first entry)
XX
DE Human PRO1124.
XX
KW Human; PRO protein; mapping.
XX
OS Homo sapiens.
XX
FN WO200116318-A2.
XX
XX
PD 08-MAR-2001..
XX
EF 24-AUG-2000; 2000WO-US23328.
XX
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 07-DEC-1999; 99US-0169495.
PR 09-DEC-1999; 99US-0170262.
PR 11-JAN-2000; 2000US-0175481.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
```

01-JUL-1998; 98US-0091360.
 PR 01-JUL-1998; 98US-0091544.
 PR 02-JUL-1998; 98US-0091478.
 PR 02-JUL-1998; 98US-0091486.
 PR 02-JUL-1998; 98US-0091519.
 PR 02-JUL-1998; 98US-0091626.
 PR 02-JUL-1998; 98US-0091628.
 PR 02-JUL-1998; 98US-0091633.
 PR 02-JUL-1998; 98US-0091646.
 PR 02-JUL-1998; 98US-0091673.
 PR 07-JUL-1998; 98US-0091978.
 PR 07-JUL-1998; 98US-0091982.
 PR 09-JUL-1998; 98US-0092182.
 PR 10-JUL-1998; 98US-0092472.
 PR 20-JUL-1998; 98US-0093339.
 PR 30-JUL-1998; 98US-0094651.
 PR 04-AUG-1998; 98US-0095282.
 PR 04-AUG-1998; 98US-0095285.
 PR 04-AUG-1998; 98US-0095301.
 PR 04-AUG-1998; 98US-0095302.
 PR 04-AUG-1998; 98US-0095318.
 PR 04-AUG-1998; 98US-0095321.
 PR 04-AUG-1998; 98US-0095325.
 PR 10-AUG-1998; 98US-0095916.
 PR 10-AUG-1998; 98US-0095929.
 PR 10-AUG-1998; 98US-0096012.
 PR 11-AUG-1998; 98US-0096143.
 PR 12-AUG-1998; 98US-0096146.
 PR 12-AUG-1998; 98US-0096329.
 PR 17-AUG-1998; 98US-0096757.
 PR 17-AUG-1998; 98US-0096766.
 PR 17-AUG-1998; 98US-0096768.
 PR 17-AUG-1998; 98US-0096773.
 PR 17-AUG-1998; 98US-0096791.
 PR 17-AUG-1998; 98US-0096867.
 PR 17-AUG-1998; 98US-0096881.
 PR 17-AUG-1998; 98US-0096894.
 PR 17-AUG-1998; 98US-0096895.
 PR 17-AUG-1998; 98US-0096897.
 PR 18-AUG-1998; 98US-0096949.
 PR 18-AUG-1998; 98US-0096950.
 PR 18-AUG-1998; 98US-0096959.
 PR 18-AUG-1998; 98US-0097061.
 PR 24-AUG-1998; 98US-0097951.
 PR 26-AUG-1998; 98US-0097952.
 PR 26-AUG-1998; 98US-0097954.
 PR 26-AUG-1998; 98US-0097955.
 PR 26-AUG-1998; 98US-0097955.
 PR 26-AUG-1998; 98US-0097971.
 PR 26-AUG-1998; 98US-0097974.
 PR 26-AUG-1998; 98US-0097978.
 PR 26-AUG-1998; 98US-0097979.
 PR 26-AUG-1998; 98US-0097986.
 PR 26-AUG-1998; 98US-0098014.
 PR 31-AUG-1998; 98US-0098525.
 PR 16-SEP-1998; 98US-0100634.
 PR 12-JAN-1999; 99US-0115565.

(GETH) GENENTECH INC.
 XX Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
 PI Wood WL, Yuan J;
 XX
 DR WPI; 2000-072883/06.
 DR N-PSDB; AAZ65095.
 XX
 XT Membrane-bound proteins and related nucleotide sequences -
 claim 12; Fig 274; 822pp; English.
 *PS
 XX

CC different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
CC are treated with P2, polynucleotides encoding P2 or antigen presenting
CC cells expressing P2 and then administered to the patient to inhibit
CC development of cancer.
XX
SQ Sequence 592 AA;

alignment_scores:

Quality: 156.50 Length: 60
Ratio: 3.402 Gaps: 1
Percent Similarity: 76.667 Percent Identity: 48.333

alignment_block:

US-09-049-696-4 x AAB11324 ..

Align seg 1/1 to: AAB11324 from: 1 to: 592

2 AAAGATGCACATCAATAAAGTAACAGGACTCTATGAAAGAGATGTGA 51
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
214 GluAsnCysIleIleSerLys.....LeuPheLysGluGlyCysTh 227
52 GTTGTCTCCAAATCCCGCAGAGGAGAGGCTTCTATATGTTGCAC 101
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
227 rPheIleTyraSnSerThrGlnAsnAlaThrAlaSerIleMetPheMetG 244
102 AACATGTTGATTCTATAGTTGAATTCGTACAGAAACAAACACCAACAAA 151
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
244 InSerLeuSerValValGluPheCysAsnAlaSerThrHisAsnGln 260
152 GAAGCTCCAAACAGCAAAATCAAAATGTC 181
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
261 GluAlaProAsnLeuGlnAsnGlnMetCys 270

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB45904

seq_documentation_block:

ID AAB45904 standard; Protein: 742 AA.

XX AAB45904;

XX 21-MAR-2001 (first entry)

XX Human tumor-associated antigen C42 protein.

XX Tumor-associated antigen; C42; human; immunogenic; cancer therapy;
KW cytostatic; immunotherapy; vaccine; lung; breast; esophagus.
XX Homo sapiens.

XX DE19924199-A1.

XX 30-NOV-2000.

XX 27-MAY-1999; 99DE-1024199.

XX 27-MAY-1999; 99DE-1024199.

XX (BOEH) BOEHRINGER INGELHEIM INT GMBH.

XX Adolf G, Heider K, Koenig U, Sommergruber W, Adolf GR, Heider K;
PI Koenig U, Sommergruber W;
XX WPI: 2001-062549/08.
XX N-PSDB; AAC82881.

XX Tumor-associated antigen C42, and related nucleic acids and antibodies,
PT useful in immunotherapy of cancer and for diagnosis -
XX Claim 1; Page 18; 40pp; German.

XX This invention describes a novel tumor-associated antigen C42 which has
XX cytostatic activity. The invention also describes (a) immunogenic protein
CC

CC fragments or peptides (I) derived from C42; (b) a pharmaceutical
CC composition for parenteral, topical, oral or local administration
CC containing at least 1 C42 or (I); (c) an isolated DNA (II) encoding C42
CC or its fragments; (d) a recombinant DNA (IIa) that contains (II); and
CC (e) antibodies (Ab) raised against C42 or (I). C42 and its fragments
CC induce a humoral immune response and, when presented by major
CC histocompatibility complex molecules, a cellular immune response. C42,
CC and its fragments, are used for immunotherapy of cancers that express C42
CC (particularly of lung, breast and esophagus), to raise specific
CC antibodies (Ab) and for diagnosis, e.g. to detect induction of an immune
CC response and for optimization of treatment. Ab are used therapeutically
CC (e.g. when conjugated to a cytotoxin) or for diagnosis or monitoring of
CC cancers that express C42. Nucleic acid (II) that encodes C42 can also be
CC used for immunotherapy and cells that express C42 as antitumor vaccines.

XX
SQ Sequence 742 AA;

alignment_scores:

Quality: 156.50 Length: 60
Ratio: 3.402 Gaps: 1
Percent Similarity: 76.667 Percent Identity: 48.333

alignment_block:

US-09-049-696-4 x AAB45904 ..

Align seg 1/1 to: AAB45904 from: 1 to: 742

2 AAAGATGCACATCAATAAAGTAACAGGACTCTATGAAAGAGATGTGA 51
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
13 GluAsnCysIleIleSerLys.....LeuPheLysGluGlyCysTh 26
52 GTTGTCTCCAAATCCCGCAGAGGAGAGGCTTCTATATGTTGCAC 101
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
26 rPheIleTyraSnSerThrGlnAsnAlaThrAlaSerIleMetPheMetG 43
102 AACATGTTGATTCTATAGTTGAATTCGTACAGAAACAAACACCAACAAA 151
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
43 InSerLeuSerValValGluPheCysAsnAlaSerThrHisAsnGln 59
152 GAAGCTCCAAACAGCAAAATCAAAATGTC 181
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
60 GluAlaProAsnLeuGlnAsnGlnMetCys 69

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT: AAY41040

seq_documentation_block:

ID AAY41040 standard; protein: 791 AA.

XX AAY41040;

XX 07-DEC-1999 (first entry)

XX Human lung tumor antigen L762P variant 2.

XX Human; lung tumor; lung cancer; T cell stimulation.

XX Homo sapiens.

XX WO9947674-A2.

XX 23-SEP-1999.

XX 17-MAR-1999; 99WO-US05798.

XX 18-MAR-1998; 98US-0040802.

XX 18-MAR-1998; 98US-0040984.

XX 27-JUL-1998; 98US-0123912.

XX 27-JUL-1998; 98US-0123933.

XX (CORI-) CORIXA CORP.

XX Read SG, Wang T;

This Page Blank (uspto)

OM of: US-09-049-696-3 to: SPTREMBL_17:* out_format : pfs

Date: Mar 30, 2002 2:46 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL-frame-n2p.model -DEV-xlp  
-Q/cgn2_1/USPTO_spool/US09049696/runat_28032002_145238_2053/app_query.fasta_1.12579  
-DB=SPTREMBL_17 -QMT-fastan -SUFFIX=rspt -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-XGAPOP=6.000 -XGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM-ext -MINLEN=0 -MAXLEN=200000000  
-USER=US09049696_cgn1_1_705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY  
-WAIT -THREADS=1
```

Search information block:

Query: US-09-049-696-3

Query length: 240

Database: SPTREMBL_17.*

Database sequences: 473505

Database length: 146272329

Search time (sec): 805.760000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
sp_human:Q9UNF6	+	442.00	932.95	1.2e-44	914	Q9unf6 homo sapiens (human). ca
sp_human:Q95151	+	439.00	926.46	2.7e-44	914	Q95151 homo sapiens (human). ca
sp_human:Q9UPC6	+	439.00	926.46	2.7e-44	914	Q9upc6 homo sapiens (human). ca
sp_mammal:Q9TUB5	+	369.00	775.02	7.4e-36	917	Q9tub5 sus scrofa (pig). epithe
sp_rodent:Q88826	+	356.00	746.94	2.7e-34	913	Q88826 mus musculus (mouse). gc
sp_rodent:Q9D726	+	356.00	746.94	2.7e-34	913	Q9d726 mus musculus (mouse). ch
sp_human:Q9UNF7	+	306.00	638.75	2.9e-28	901	Q9unf7 homo sapiens (human). ca
sp_rodent:Q88860	+	266.00	552.37	1.9e-23	902	Q88860 mus musculus (mouse). ch
sp_rodent:Q9GX15	+	265.00	552.36	1.9e-23	902	Q9gx15 mus musculus (mouse). ca
sp_human:Q9Y6N3	+	265.00	560.37	2.3e-23	902	Q9y6n3 homo sapiens (human). cl
sp_rodent:Q9R070	+	264.00	548.03	3.3e-23	902	Q9r070 mus musculus (mouse). ca
sp_rodent:Q9EQ84	+	264.00	548.03	3.3e-23	902	Q9eq84 mus musculus (mouse). en
sp_mammal:Q18744	+	255.00	536.54	3.8e-22	342	Q18744 bos taurus (bovine). lu
sp_mammal:Q18743	+	255.00	529.61	4.0e-22	794	Q18743 bos taurus (bovine). lu
sp_mammal:Q18743	+	255.00	529.35	4.0e-22	820	Q18743 bos taurus (bovine). lu
sp_mammal:Q18741	+	255.00	528.54	4.0e-22	905	Q18741 bos taurus (bovine). lu
sp_human:Q9Y6N2	+	239.50	494.67	3.0e-20	943	Q9y6n2 homo sapiens (human). ca
sp_human:Q9UC09	+	230.50	475.20	3.6e-19	943	Q9uc09 homo sapiens (human). ch
sp_bacteria:Q9I188	+	73.50	140.12	2.88	544	Q9i188 pseudomonas aeruginosa. mc
sp_plant:Q80780	+	73.50	140.25	3.29	470	Q80780 arabidopsis thaliana (mc
sp_invertebrate:Q23961	+	67.50	121.91	15.78	156	Q23961 drosophila kitumensis
sp_plant:Q9C9L3	+	66.50	124.73	20.14	1028	Q9c9l3 arabidopsis thaliana (m
sp_bacteria:Q86842	+	66.50	124.73	20.14	561	Q86842 streptomyces coelicolor. m
sp_invertebrate:Q17430	+	66.50	121.45	20.59	836	Q17430 caenorhabditis elegans
sp_organelle:Q9MES6	+	65.50	121.93	26.69	606	Q9mes6 physeter catodon (specm
sp_invertebrate:Q9W1M0	+	65.50	118.93	27.23	873	Q9w1m0 drosophila melanogaste
sp_invertebrate:Q9W4S6	+	65.00	131.31	28.58	170	Q9w4s6 drosophila melanogaste
sp_plant:Q39810	+	65.00	124.50	29.92	389	Q39810 glycine max (soybean). a
sp_bacteria:Q9FAX4	+	65.00	122.75	30.27	481	Q9fax4 capnocytophaga ochracea. m
sp_organelle:Q9B2G0	+	64.50	122.05	34.69	459	Q9b2g0 isocodon macrourus (short
sp_invertebrate:Q24573	+	64.50	116.51	36.00	901	Q24573 drosophila melanogaste
sp_bacteria:Q9IOM9	+	64.00	123.76	39.11	327	Q9iom9 pseudomonas aeruginosa. m
sp_human:Q9NV84	+	64.00	118.17	40.60	645	Q9nv84 homo sapiens (human). co
sp_human:Q9B064	+	64.00	118.17	40.60	645	Q9bq64 homo sapiens (human). hy
sp_human:Q9H869	+	64.00	116.44	41.08	796	Q9h869 homo sapiens (human). co
sp_human:Q9BQ11	+	64.00	116.05	41.19	835	Q9bq11 homo sapiens (human). hy
sp_bacteria:Q9PK94	+	63.50	128.67	43.16	158	Q9pk94 chlamydia muridarum. hyp
sp_plant:Q82568	+	63.50	121.25	45.36	389	Q82568 arabidopsis thaliana (mc
sp_invertebrate:Q18552	+	63.50	119.29	45.96	494	Q18552 drosophila pseudobscu
sp_bacteria:Q9K437	+	63.50	116.50	46.83	693	Q9k437 streptomyces coelicolor. l

```
sp_bacteria:Q9AJF8 + 63.50 116.30 46.89 710 ! Q9ajf8 clostridium thermocel  
sp_plant:Q9SKA9 - 63.50 116.02 46.98 735 ! Q9ska9 arabidopsis thaliana  
sp_invertebrate:Q02462 - 63.00 125.90 50.14 194 ! Q02462 pedetontus saltator  
sp_bacteria:Q9LBG4 + 63.00 122.70 51.23 286 ! Q9lbg4 amycolatopsis sp. cso  
sp_bacteria:Q46377 + 63.00 119.82 52.23 406 ! Q46377 comamonas testosteron  
seq_name: sp_human:Q9UNF6  
seq_documentation_block:  
ID Q9UNF6 PRELIMINARY; PRT; 914 AA.  
AC Q9UNF6;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 1.  
GN CACCL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI_TaxID=9606;  
PN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SMALL INTESTINE, AND COLON;  
RX MEDLINE=99364503; PubMed=10437792;  
RA Agnel M., Vermaat F., Culouscou J.M.;  
RT Identification of three novel members of the calcium-dependent  
RT chloride channel (CACCL) family predominantly expressed in the  
RT digestive tract and trachea.;  
RL FEBS Lett. 455:295-301(1999).  
DR EMBL; AF127036; AAD25487.1; ..  
DR InterPro: IPR000131; ATPase_gamma.  
DR InterPro: IPR002035; VWFA.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; P500153; ATPASE_GAMMA; UNKNOWN_1.  
DR PROSITE; P50234; VWFA; 1.  
SQ SEQUENCE 914 AA; 100202 MW; B58B451D97887D6B CRC64;
```

```
alignment_scores:  
Quality: 442.00 Length: 79  
Ratio: 5.595 Gaps: 0  
Percent similarity: 100.00 Percent identity: 100.00  
alignment_block:  
US-09-049-696-3 x Q9UNF6 ..  
! Align 'seg 1/1' to: Q9UNF6 from: 1 to: 914  
.. 3 AATCCTGATGTTCTGGTGGCTGAGTCTACTCTCCAGGTATGATGAACC 52  
||||| 101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117  
.. 53 CTACACTGAGCAGATGGCAACTGTGGAGAGAGGGTGAAGGATCCACC 102  
||||| 117 cTyrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134  
.. 103 TCACCTCTGATTTCAATTCAGCAAGAAAAAGTTAGCTGAATATGGACCAA 152  
||||| 134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150  
.. 153 GGTAGGCATTTGTCATGAGTGGGCTCATCTACGATGGGAGTATTGA 202  
||||| 151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167  
.. 203 CGACTACAATAATGATCAGAAATTTCTACTTATCCAAT 239  
||||| 167 pGluTyrAsnAspGluLysPheTyrLeuSerAsn 179  
seq_name: sp_human:Q95151  
seq_documentation_block:  
ID Q95151 PRELIMINARY; PRT; 914 AA.  
AC Q95151;
```

```
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIDIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
ON NCBI_TaxID=9606;
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Eible R.C., Ji H.L., Schreuer K.D., Fuller C.M.,
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
of human CLCA1, the first human member of the family of Ca2+-activated
Cl-channel proteins."
RL Genomics 54:200-214(1998).
DR EMBL; AF039400; AAC95428.1; -.
DR InterPro; IPR000131; ATPase_gamma.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 914 AA; 100174 MW; 7066440C46557FA3 CRC64;

alignment_scores:
  Quality: 439.00      Length: 79
  Ratio: 5.557        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.734

alignment_block:
US-09-049-696-3 x Q95151 ..
Align seg 1/1 to: Q95151 from: 1 to: 914

3  AATGCTGATGTTCTGCTGAGTCTACTCTCCAGGTAATGATGAACC 52
|||||
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117

53  CTACACTGACGAGTGGCACTGTGGAGAGAGGTGAAGATCCACC 102
|||||
117 oTyrThrGluGlnMetGlyAsnCysGlyGluGlyGluArgGluHisL 134

103 TCACCTCTGATTTCATTCAGGAGAAAAGTTAGCTGAATATGGACACAA 152
|||||
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrglyProGln 150

153 GGTAGGGCATTGTCCTCATGAGTGGCTCATCTACGATGGGAGTATTGA 202
|||||
151 GlyLysAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167

203 CGAGTACAATAATGATGAGAAATTTACTATTCCAAT 239
|||||
167 pGluTyraAsnAspGluLysPheTyrlSerAsn 179

seq_name: sp_human:Q9UPC6

seq_documentation_block:
ID Q9UPC6 PRELIMINARY; PRT; 914 AA.
AC Q9UPC6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIDIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
ON NCBI_TaxID=9606;
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Eible R.C., Ji H.L., Schreuer K.D., Fuller C.M.,
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
of human CLCA1, the first human member of the family of Ca2+-activated
Cl-channel proteins."
RL Genomics 54:200-214(1998).
DR EMBL; AF039400; AAC95428.1; -.
DR InterPro; IPR000131; ATPase_gamma.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 914 AA; 100174 MW; 7066440C46557FA3 CRC64;

alignment_scores:
  Quality: 439.00      Length: 79
  Ratio: 5.557        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.734

alignment_block:
US-09-049-696-3 x Q95151 ..
Align seg 1/1 to: Q95151 from: 1 to: 914

3  AATGCTGATGTTCTGCTGAGTCTACTCTCCAGGTAATGATGAACC 52
|||||
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117

53  CTACACTGACGAGTGGCACTGTGGAGAGAGGTGAAGATCCACC 102
|||||
117 oTyrThrGluGlnMetGlyAsnCysGlyGluGlyGluArgGluHisL 134

103 TCACCTCTGATTTCATTCAGGAGAAAAGTTAGCTGAATATGGACACAA 152
|||||
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrglyProGln 150

153 GGTAGGGCATTGTCCTCATGAGTGGCTCATCTACGATGGGAGTATTGA 202
|||||
151 GlyLysAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167

203 CGAGTACAATAATGATGAGAAATTTACTATTCCAAT 239
|||||
167 pGluTyraAsnAspGluLysPheTyrlSerAsn 179

seq_name: sp_human:Q9UPC6

seq_documentation_block:
ID Q9UPC6 PRELIMINARY; PRT; 914 AA.
AC Q9UPC6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIDIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
ON NCBI_TaxID=9606;
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Eible R.C., Ji H.L., Schreuer K.D., Fuller C.M.,
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
of human CLCA1, the first human member of the family of Ca2+-activated
Cl-channel proteins."
RL Genomics 54:200-214(1998).
DR EMBL; AF039400; AAC95428.1; -.
DR InterPro; IPR000131; ATPase_gamma.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 914 AA; 100174 MW; 7066440C46557FA3 CRC64;
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Eible R.C., Ji H.L., Schreuer K.D., Fuller C.M.,
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
of human CLCA1, the first human member of the family of Ca2+-activated
Cl-channel proteins."
RL Genomics 54:200-214(1998).
DR EMBL; AF039401; AAC95429.1; -.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 914 AA; 100160 MW; 7761362B66752AA3 CRC64;

alignment_scores:
  Quality: 439.00      Length: 79
  Ratio: 5.557        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.734

alignment_block:
US-09-049-696-3 x Q9UPC6 ..
Align seg 1/1 to: Q9UPC6 from: 1 to: 914

3  AATGCTGATGTTCTGCTGAGTCTACTCTCCAGGTAATGATGAACC 52
|||||
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117

53  CTACACTGACGAGTGGCACTGTGGAGAGAGGTGAAGATCCACC 102
|||||
117 oTyrThrGluGlnMetGlyAsnCysGlyGluGlyGluArgGluHisL 134

103 TCACCTCTGATTTCATTCAGGAGAAAAGTTAGCTGAATATGGACACAA 152
|||||
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrglyProGln 150

153 GGTAGGGCATTGTCCTCATGAGTGGCTCATCTACGATGGGAGTATTGA 202
|||||
151 GlyLysAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167

203 CGAGTACAATAATGATGAGAAATTTACTATTCCAAT 239
|||||
167 pGluTyraAsnAspGluLysPheTyrlSerAsn 179

seq_name: sp_mammal:Q9TUB5

seq_documentation_block:
ID Q9TUB5 PRELIMINARY; PRT; 917 AA.
AC Q9TUB5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN.
GN AEC.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
ON NCBI_TaxID=9823;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ILEAL MUCOSA;
RA Gaspar K.J., Gabriel S.E., Racette K.J., Forsyth G.W.;
RT "A cDNA involved in porcine exocrine chloride conductance."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF095384; AAF00077.1; -.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
```

SQ SEQUENCE 917 AA; 100735 MW; 649F41944F7EAD19 CRC64;

alignment_scores:
Quality: 369.00 Length: 79
Ratio: 4.920 Gaps: 0
Percent Similarity: 94.937 Percent Identity: 78.481

alignment_block:

US-09-049-696-3 x Q9TUB5

Align seg 1/1 to: Q9TUB5 from: 1 to: 917

```

3 AATGCTGATGTTCTGTTGCTGAGTCTACTCTCCAGGTAATGATGAACC 52
|||||
101 AsnAlaAspValValValThrGluProAsnProGluAsnAspGlyPr 117
|||||
53 CTACACTGACGACATGGCAACTGTGAGAGAGAGGTTGAAGGATCCACC 102
|||||
117 oTyThrGluGlnMetGlyAsnCysGlyGluLysGlyGluLysIleTyrP 134
|||||
103 TCACCTCTGATTTCAATGCAGGAAAAAGTTAGCTGAATATGACACCAA 152
|||||
134 heThrProAspPheValAlaGlyLysLysValLeuGlnTyrGlyProGln 150
|||||
153 GGTAGGCATTTGTCATGAGTGGGCTCATCTACGATGGGAGTATTGTA 202
|||||
151 GlyArgValPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
|||||
203 CGAGTACAATAATGATGAGAAATTTCTACTTATCCAT 239
|||||
167 nGluTyrAsnAsnGluGlnLysPheTyrLeuSerAsn 179
|||||

```

seq_name: sp_rodent:O88826

seq_documentation_block:

```

AC O88826; PRELIMINARY; PRT; 913 AA.
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GOB-5 PROTEIN.
GN CLCA3 OR GOB-5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RA Komiyama T., Tanigawa Y., Hirohashi S.;
RT "Cloning of a gene, gob-5, which is expressed in intestinal goblet
cells in mice."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017156; BAA33743.1; -.
DR MGD; MGI:1346342; Clca3
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 913 AA; 100770 MW; A7FA2F9E1089806D CRC64;

```

alignment_scores:
Quality: 356.00 Length: 78
Ratio: 4.877 Gaps: 0
Percent Similarity: 93.590 Percent Identity: 79.487

alignment_block:

US-09-049-696-3 x O88826

Align seg 1/1 to: O88826 from: 1 to: 913

```

3 AATGCTGATGTTCTGTTGCTGAGTCTACTCTCCAGGTAATGATGAACC 52
|||||
101 AsnAlaAspValValValThrThrSerProLeuGlyAsnAspGluPr 117
|||||
53 CTACACTGACGACATGGCAACTGTGAGAGAGAGGTTGAAGGATCCACC 102
|||||
117 oTyThrGluHisIleGlyAlaCysGlyGluLysGlyIleArgIleHisL 134
|||||
103 TCACCTCTGATTTCAATGCAGGAAAAAGTTAGCTGAATATGACACCAA 152
|||||
134 eutrProAspPheLeuAlaGlyLysLysLeuThrGlnTyrGlyProGln 150
|||||
153 GGTAGGCATTTGTCATGAGTGGGCTCATCTACGATGGGAGTATTGTA 202
|||||
151 AspArgThrPheValHisGluTrpAlaHisPheArgTrpGlyValPheAs 167
|||||
203 CGAGTACAATAATGATGAGAAATTTCTACTTATCC 236
|||||
167 nGluTyrAsnAsnAspGluLysPheTyrLeuSer 178
|||||

```

seq_name: sp_rodent:Q9D7Z6

seq_documentation_block:

```

AC Q9D7Z6; PRELIMINARY; PRT; 913 AA.
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CHLORIDE CHANNEL CALCIUM ACTIVATED 3.
GN CLCA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=STOMACH;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Willing L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK008659; BAB25815.1; -.
DR MGD; MGI:1346342; Clca3.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 913 AA; 100111 MW; 491E8584B06D9A89 CRC64;

```

alignment_scores:

Quality: 356.00 Length: 78
Ratio: 4.877 Gaps: 0
Percent Similarity: 93.590 Percent Identity: 79.487

alignment_block:

US-09-049-696-3 x Q9D726 ..

Align seg 1/1 to: Q9D726 from: 1 to: 913

```
3  AATGCTGATGTTCTGCTGCTGAGTCTACTCTCCAGGTAATGATGAACC 52
   |||.....|
101 AsnAlaAspValLeuValSerThrThrSerProLeuGlyAsnAspGluPr 117
   |||.....|
53  CTACACTGACGAGTGGCAACTGTGGAGAGAGGTGAAGGATCCACC 102
   |||.....|
117 oTyrThrGluHisIleGlyAlaCysGlyGlyLysGlyIleArgIleHisL 134
   |||.....|
103 TCACCTCTGATTTCATTCGAGGAAAAAGTTAGCTGAATATGACACACAA 152
   |||.....|
134 euThrProAspPheLeuAlaGlyLysLysLeuThrGlnTyrGlyProGln 150
   |||.....|
153 GGTAGGCGATTGTCCTGCTGAGTGGGCTCATCTACGATGGGAGTATTGA 202
   |||.....|
151 AspArgThrPheValHisGluTrpAlaHisPheArgTrpGlyValPheAs 167
   |||.....|
203 CGAGTACAATAATGATGAGAAATTCCTATCC 236
   |||.....|
167 nGluTyAsnAsnAspGluLysPheTyrLeuSer 178
```

seq_name: sp_human:Q9UNF7

seq_documentation_block:

ID Q9UNF7 PRELIMINARY; PRT; 917 AA.

```
AC Q9UNF7
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 2.
GN CACCC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RX MEDLINE=9364503; PubMed=10437792;
RA Agnel M., Vermet T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
RT chloride channel (CaCC) family predominantly expressed in the
RT digestive tract and trachea.";
RL FEBS Lett. 455:295-301(1999).
DR EMBL: AF127035; AAD48398.1; -.
DR InterPro: IPR002035; VWFA.
DR Pfam: PF00092; vwa; 1.
DR DR PROSITE: PS50234; VWFA; 1.
DR SMART: SM00327; VWA; 1.
SQ SEQUENCE 917 AA; 101153 MW; EA01090D781BEB95 CRC64;
```

alignment_scores:

Quality: 306.00 Length: 76
Ratio: 4.708 Gaps: 0
Percent Similarity: 85.526 Percent Identity: 69.737

alignment_block:

US-09-049-696-3 x Q9UNF7

Align seg 1/1 to: Q9UNF7 from: 1 to: 917

```
3  AATGCTGATGTTCTGCTGCTGAGTCTACTCTCCAGGTAATGATGAACC 52
   |||.....|
100 HisAlaAspValIleValAlaProThrLeuProGlyArgAspGluPr 116
   |||.....|
53  CTACACTGACGAGTGGCAACTGTGGAGAGAGGTGAAGGATCCACC 102
   |||.....|
```

```
116 oTyrThrLysGlnPheThrGluCysGlyGlyLysGlyGluTyrIleHisP 133
103 TCACCTCTGATTTCATTCGACGAAAAAGTTAGCTGAATATGACACCAA 152
   |||.....|
133 heThrProAspLeuLeuGlyLysGlnAsnGluTyrGlyProPro 149
   |||.....|
153 GGTAGGCGATTGTCCTGCTGAGTGGGCTCATCTACGATGGGAGTATTGA 202
   |||.....|
150 GlyLysLeuPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 166
   |||.....|
203 CGAGTACAATAATGATGAGAAATTCCTAC 230
   |||.....|
166 pGluTyAsnGluAspGlnProPheTyr 175
```

seq_name: sp_rodent:O88860

seq_documentation_block:

ID O88860 PRELIMINARY; PRT; 901 AA.

```
AC O88860
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CHLORIDE CHANNEL CACC.
GN CLCAL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
PA Romio L., Musante L., Cinti R., Moran O., Seri M., Galletta L.J.V.;
RT "Characterization of a murine gene homologous to the bovine CaCC
RT chloride channel.";
KL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF052746; AAC35003.1; -.
DR MGD: MGI:1316732; Clcal.
DR InterPro: IPR002035; VWFA.
DR SMART: SM00327; VWA; 1.
DR PROSITE: PS50234; VWFA; 1.
SQ SEQUENCE 901 AA; 100038 MW; 13DF068BF33737AAB CRC64;
```

alignment_scores:

Quality: 266.00 Length: 77
Ratio: 4.290 Gaps: 0
Percent Similarity: 80.519 Percent Identity: 59.740

alignment_block:

US-09-049-696-3 x O88860 ..

Align seg 1/1 to: O88860 from: 1 to: 901

```
6  GCTGATGTTCTGCTGCTGAGTCTACTCTCCAGGTAATGATGAACCCTA 55
   |||.....|
101 AlaAspValIleValAlaAspProHisLeuGlnHisGlyAspAspProTy 117
   |||.....|
56  CACTGACGAGATGGGCAACTGTGGAGAGAGGTGAAGGATCCACCTCA 105
   |||.....|
117 rThrLeuGlnTyrGlyGlnCysGlyAspArgGlyGlnTyrIleHisPheT 134
   |||.....|
106 CTCCTGATTTCATTCGAGGAAAAAGTTAGCTGAATATGACACCAAGGT 155
   |||.....|
134 hrProAsnPheLeuLeuThrAspAsnLeuArgIleTyrGlyProArgGly 150
   |||.....|
156 AGGCGATTGTCCTGCTGAGTGGGCTCATCTACGATGGGAGTATTGACGA 205
   |||.....|
151 ArgValPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAspG 167
   |||.....|
206 CTACATAATGATGAGAAATTCCTATCC 236
   |||.....|
167 uTyAsnValAspGlnProPheTyrMetSer 177
```

seq_name: sp_rodent:O90X15

OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=WHOLE LUNG;	
RX	MEDLINE=99041980; PubMed=9822685;	
RA	Gandhi R., Elbie R.C., Gruber A.D., Schr�ur K.D., Ji H.-L.,	
RA	Fuller C.M., Pauli B.U.	
RT	"Molecular and functional characterization of a calcium-sensitive	
RT	chloride channel from mouse lung."	
RL	J. Biol. Chem. 273:32096-32101(1998).	
DR	ENBL; AF047838; AAC79982.1; "	
DR	MGD; MGI:1316732; Cical.	
DR	InterPro: IPR002035; vWFA.	
DR	SMART: SM00327; vWA; 1.	
DR	PROSITE; PS50234; vWFA; 1.	
SQ	SEQUENCE 902 AA; 100107 MW; D9A915E698EA2D85 CRC64;	

Align,seg 1/1 to: Q9Y6N3 from: 1 to: 262

GN GUCA2 OR CACC.
 OS Mus musculus (Mouse).
 CS Eukaryota; Metazoa;
 OC Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia;
 OX Sciurognathi; Muridae; Murinae; Mus.
 QX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 NC TISSUE=MAMMARY.

DR	EK3L; AFI08501; AAF12731.1; -	
DR	MCD; MGI:1931471; Clc62.	
DR	InterPro; IPR02035; vWFA.	
DR	SMART; SM00327; vWA; 1.	
DR	ProSITE; PS02034; vWFA; 1.	
SQ	SEQUENCE 902 AA; 99866 MW;	0FB746CF5EB3D07E C

alignment_scores:

Quality:	264.00	Length:	77
Ratio:	4.258	Gaps:	0
Percent Similarity:	80.519	Percent Identity:	59.740

alignment block:

US-09-049-696-3 x Q9R070

Align seg 1/1 to: Q9R070 from: 1 to: 902

6 GCTGATGTTCTGGTTCAGTGCTACTCCTCCAGGTAATGATGAACCTA 55
|||||:::|||||::
101 ALaAsnValIleValAlaAsnProHisLeuClnhisClValAsnAspMetV 117

101 AARASPVAVLEAVAVAGASPRFRIHSLEUGLNHISGLYASPASPRQY 111
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117 rThrLeuGlnTyrGlyGlnCysGlyASAPArqGlyGlnTyrIleHisPheT 134
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156 CACTGAGCAGATGGCACTGTGGAGACAGAGGTGAAGGATCCACTCA 105
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900

106 CTCCTGATTTCATTCGCAGGAAAAAGTTAGCTGAATATGGACCACCAAGGT 155
|||||::|||::: |||::: |||::: |||::: |||
134 hrProAsnPheLeuLeuthrAspAsnLeuAraIletrVGlyproArqGly 150

156 AGGCATTTGTCCATGAGTGGGTCATCTACGATGGGAGTATTTGACGA 205
 151 ArgValPheValHisGluTrpAlaHisIleuAcrTropGlyValPheAspGly 167

206 GTACAATAATGATGAGAAATTCCTATTATCC 236
 ||||| |||: |||||:|
 167 uTvAsnVaJsoAraGPropheTvTleser 177

```
seq name: sp_rodent:Q9EOR4
100  arjmsvunspgfrormeyllleber 177
```

```
seq_documentation_block:
ID_Q9EQR4
PRELIMINARY:
PRT: 902 AA.
```

AC	Q9E0R4:	
DT	01-MAR-2001	(TREMBlrel. 16, Created)
DT	01-MAR-2001	(TREMBlrel. 16, Last sequence update)
DT	01-JUN-2001	(TREMBlrel. 17, Last annotation update)
DE	ENDOTOXIAL	CHLORIDE CHANNEL.

GN CLCA2.

OS *Mus musculus* (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus

```

OX NCBI_TaxID=10090;
RN {1}

```

RP SEQUENCE FROM N.A.
RC STRAIN=C57/BL6;

RA Elble R.C., Pauli B.;
RT "A murine endothelial
homolog of Lu-ECAM-1 associated with

Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
EMBL Accession No. AF047626

DR EMBL; AF115852; AAG4/626.1; -
DR InterPro; IPR003961; FN_III.
DB InterPro; IPR002035; vJEA

```
DR INTCFIO; IFR002035; VWFA;
DR SMART; SM00060; FN3; 1.
DR SMART; SM00327; VWFA; 1
```

DR SHARI, SH000327, VWB, I.
DR PROSITE; PS50234; VWFA; 1.
SO SEQUENCE 902 AA: 99855 MW: F2BA57DAA5D4AAAD CRC64.

[illegible]

alignment_scores:	
Quality:	264.00
Length:	77

Ratio: 4.258 Gaps: 0
Percent Similarity: 80.519 Percent Identity: 59.740

alignment_block:

US-09-049-696-3 x Q9EQR4

Align seg 1/1 to: Q9EQR4 from: 1 to: 902

[illegible]

101 AlaAspValIleValAlaAspProHisLeuGlnHisGlyAspAspProty 117

This Page Blank (uspto)

OM of: US-09-049-696-3 to: SwissProt_39:* out_format : pfs

Date: Mar 30, 2002 2:51 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+ntp.model -DEV=xlp
-O/cygn2_1/USPRO.spool/US09049696/runat_28032002_145238_2085/app_query.fasta_1.12579
-DB=SwissProt_39 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696_ECGN1_1.165 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-049-696-3
Query length: 240
Database: SwissProt_39:*
Database sequences: 100059
Database length: 36664827
Search time (sec): 306.030000
```

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
SwissProt_39:ECIC_BOVIN	+	265.00	546.73	9.8e-24	903 ! P54281 bos taurus (bovine).
SwissProt_39:YPIP_LACDL	+	71.50	144.59	1.28	173 ! P46543 lactobacillus delbrue
SwissProt_39:EX5B_MCTCT	+	68.50	122.19	3.58	1094 ! P96920 mycobacterium tuber
SwissProt_39:SNDH_ACELI	+	67.50	127.74	4.29	449 ! Q44091 acetobacter liquefac
SwissProt_39:GUNV_ERWCA	+	67.00	125.65	4.98	505 ! Q47096 erwinia carotovora. e
SwissProt_39:NTSB_LACLA	+	66.00	117.65	7.07	993 ! P20103 lactococcus lactis (s
SwissProt_39:PSPA_HUMAN	+	63.00	123.19	13.92	248 ! P07714 homo sapiens (human).
SwissProt_39:PGK_APLCA	+	63.00	118.80	14.70	412 ! P61471 aplysia californica (c
SwissProt_39:PGK_OPIA	+	63.00	118.74	14.72	415 ! P50311 opisthorchis sinensis
SwissProt_39:CDRI_CANAL	-	63.00	107.62	16.92	1501 ! P43071 candida albicans (ye
SwissProt_39:Y085_YEATJ	+	62.50	118.58	16.70	373 ! Q57550 methanococcus jannasch
SwissProt_39:YN87_YEAST	+	62.00	122.39	18.03	212 ! P33719 saccharomyces cerevis
SwissProt_39:MCAS_MYCBO	+	62.00	102.53	23.14	2110 ! Q02251 mycobacterium bovis.
SwissProt_39:ATH1_YEAST	+	61.50	106.25	25.01	1211 ! P48016 saccharomyces cerevi
SwissProt_39:ZEP1_MOUSE	+	61.50	99.36	27.27	2688 ! Q03172 mus musculus (mouse)
SwissProt_39:GLI4_HUMAN	+	61.00	119.80	23.89	223 ! P10075 homo sapiens (human).
SwissProt_39:TPIS_MORSP	+	61.00	118.18	24.38	269 ! Q10893 moraxella sp. trioseph
SwissProt_39:CHI3_CANAL	+	61.00	111.74	26.44	567 ! P40954 candida albicans (yea
SwissProt_39:WNSB_HUMAN	-	60.50	114.80	28.81	351 ! Q93098 homo sapiens (human).
SwissProt_39:GUN1_CLOTH	+	60.50	106.87	31.83	879 ! Q02934 clostridium thermocel
SwissProt_39:MSRE_HUMAN	+	60.00	111.56	33.99	451 ! P21757 homo sapiens (human)
SwissProt_39:GUNZ_CLOS	+	60.00	104.80	37.00	986 ! P23659 clostridium stercorari
SwissProt_39:YHL4_HCMVA	+	59.50	118.82	35.14	172 ! P09698 human cytomegalovirus
SwissProt_39:CHIS_NOCSP	+	59.50	114.67	37.02	278 ! P48846 nocardioides sp. (str
SwissProt_39:WNSB_MOUSE	-	59.50	112.68	37.96	350 ! Q57422 mus musculus (mouse).
SwissProt_39:FA19_MOUSE	+	59.50	104.31	42.16	921 ! Q05722 mus musculus (mouse).
SwissProt_39:FA13_CANAL	+	59.50	97.45	45.96	2037 ! P34731 c fatty acid synthas
SwissProt_39:ALG3_YEAST	+	59.00	109.28	44.87	458 ! P38179 saccharomyces cerevis
SwissProt_39:CAL4_DROME	+	59.00	97.57	51.97	1775 ! P08120 drosophila melanogast
SwissProt_39:VMT2_TNABC	-	58.50	115.58	46.95	195 ! P13881 influenza b virus (st
SwissProt_39:VMT2_TNABD	-	58.50	115.58	46.95	195 ! P13882 influenza b virus (st
SwissProt_39:VMT2_TNBSI	-	58.50	115.58	46.95	195 ! P08383 influenza b virus (st
SwissProt_39:SA22_HUMAN	+	58.50	113.30	48.31	254 ! P04226 homo sapiens (human).
SwissProt_39:SSUC_ECOLI	+	58.50	113.00	48.50	263 ! P75851 escherichia coli. put
SwissProt_39:PP12_MOUSE	+	58.50	112.77	48.63	270 ! P53811 mus musculus (mouse).
SwissProt_39:PP12_RAT	+	58.50	112.77	48.63	270 ! P53812 rattus norvegicus (rat)
SwissProt_39:HMS_DROME	-	58.50	107.49	51.97	497 ! P18488 drosophila melanogast
SwissProt_39:SIU7_YEAST	-	58.00	108.69	57.97	382 ! Q02775 saccharomyces cerevis
SwissProt_39:YAFK_RHISN	-	58.00	107.73	58.68	427 ! P55449 rhizobium sp. (strain
SwissProt_39:NU4M_DASNO	+	58.00	107.11	59.14	459 ! Q21334 dasyatis novemcinctus

```
SwissProt_39:NU4M_DIDMA + 58.00 106.83 59.35 474 ! P41308 didelphis marsupia
SwissProt_39:ENV_HVIMN + 58.00 101.72 63.28 856 ! P05877 homo immunodeficie
SwissProt_39:DYHC_HUMAN + 58.00 101.31 63.60 897 ! Q14204 homo sapiens (huma
SwissProt_39:CA16_CHICK + 58.00 100.21 64.49 1019 ! P20785 gallus gallus (ch
SwissProt_39:TRAA_RHISN + 58.00 99.54 65.04 1102 ! P55418 rhizobium sp. (st
```

seq_name: SwissProt_39:ECIC_BOVIN

seq_documentation_block:

```
ID ECIC_BOVIN STANDARD; PRT; 903 AA.
AC P54281;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN (CALCIUM-ACTIVATED CHLORIDE
DE CHANNEL).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Trachea;
RX MEDLINE=96125078; PubMed=8537359;
RA Cunningham S.A., Awayda M.S., Bublief J.K., Ismailov I.I.,
RA Arrate M.P., Berdiev B.K., Benos D.J., Fuller C.M.;
RT "Cloning of an epithelial chloride channel from bovine trachea.";
KL J. Biol. Chem. 270:31016-31026(1995).
CC !- FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE
CC !- SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME;
CC !- MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND
CC !- TRANSEPITHELIAL TRANSPORT.
CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC !- TISSUE SPECIFICITY: TRACHEA.
CC !- PTM: PHOSPHORYLATED BY CAM-KINASE II.
CC !- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

```
EMBL; U36445; AAC48511.1; -.
InterPro: IPR002035; VWFA.
SMART; SM00327; VWFA; 1.
PROSITE; PS50234; VWFA; 1.
Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
Calcium channel; Phosphorylation; Glycoprotein.
TRANSMEM 7 27 POTENTIAL.
TRANSMEM 331 351 POTENTIAL.
TRANSMEM 617 637 POTENTIAL.
TRANSMEM 883 903 POTENTIAL.
DOMAIN 308 476 VWFA.
FT CARBOHYD 75 75 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 688 688 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 811 811 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 816 816 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 842 842 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 857 857 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 903 AA; 100305 MW; 9742AB1590908AEC CRC64;
```

alignment_scores: Quality: 265.00 Length: 77


```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X68307; CAA48381.1; -
DR EMBL; L16226; AAA25190.1; -
DR EMBL; M65089; AAA73039.1; -
DR EMBL; J04057; AAA88607.1; -
DR EMBL; D10768; BAA01599.1; -
DR PIR; C31915; C31915.
DR PIR; B43743; B43743.
DR PIR; C48951; C48951.
KW Transport; Transmembrane.
FT TRANSMEM 838 851
FT CONFLICT 19 19
FT CONFLICT 656 656
FT CONFLICT 841 852
FT CONFLICT 895 895
SQ SEQUENCE 993 AA; 117501 MW; 0027053BEAE71E2D CRC64;

alignment_scores:
  Quality: 66.00      Length: 74
  Ratio: 1.692       Gaps: 3
  Percent Similarity: 52.703      Percent Identity: 28.378

alignment_block:
US-09-049-696-3 x NISB_LACLA ..
Align seg 1/1 to: NISB_LACLA from: 1 to: 993

60 GAGCAGTGGCGCACTGTGAGAGAGGGTGAAGGATCCACCTCACTCC 109
|||||: ||| |||||||||: |||:
687 Gluasnilleleasnlysglyglulysglyargvalalaspvalvalva 703
|||||: ||| |||||||||: |||:
110 TGATTTCATTGCGAGGAAAAGTAGCTGAATGGACCAACAAGTAGGG 159
|||||: ||| |||
703 lPropheilleargThrArgalalaLeu.....GlyAsnGluGlyArg 717

160 CATTTCGTC..... 167
|||||:
717 laPheilleArgGluLysArgValSerValGluArgArgGluLysLeuPro 733

168 ...CATGAGTGGCGTCATCTACGATGGGAGTATTTCACGAGTACATAA 214
|||||: ||| |||: |||: |||:
734 PheAsnGluTrpLeuTyrlsLeuLys.....LeuTyrlsSerileAsnAr 748

215 TGATGAGAAATTCATTATTC 236
|||||: ||| |||
748 gGlnAsnGluPheLeuLeuSer 755

seq_name: SwissProt_39:PSPA_HUMAN

seq_documentation_block:
ID PSPA_HUMAN STANDARD; PRT; 248 AA.
AC P07714;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PULMONARY SURFACTANT-ASSOCIATED PROTEIN A PRECURSOR (SP-A) (PSP-A)
DE (PSPAP) (ALVEOLAR PROTEINOSIS PROTEIN) (35 KDA PULMONARY SURFACTANT-
DE ASSOCIATED PROTEIN).
GN SFTPA1 OR SFTPA OR SFTP1 OR PSPAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86250832; PubMed=3755136;
RA Floros J., Steinbrink R., Jacobs K., Phelps D., Kriz R., Recny M.,
RA Sultzman L., Jones S., Tausch H.W., Frank H.A., Fritsch E.F.;
RT "Isolation and characterization of cDNA clones for the 35-kDa
RT pulmonary surfactant-associated protein.";
RL J. Biol. Chem. 261:9029-9033(1986).
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=86014366; PubMed=2995821;
RA White R.T., Damm D., Miller J., Spratt K., Schilling J., Hawgood S.,
RA Bepson B., Coddell B.;
RT "Isolation and characterization of the human pulmonary surfactant
RT apoprotein gene.";
RL Nature 317:361-363(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92198680; PubMed=1372511;
RA Kalyal S.L., Singh G., Locker J.L.;
RT "Characterization of a second human pulmonary surfactant-associated
RT protein SP-A gene.";
RL Am. J. Respir. Cell Mol. Biol. 6:446-452(1992).
CC FUNCTION: IN PRESENCE OF CALCIUM IONS, PSAP BINDS TO SURFACTANT
CC PHOSPHOLIPIDS AND CONTRIBUTES TO LOWER THE SURFACE TENSION AT THE
CC AIR-LIQUID INTERFACE IN THE ALVEOLI OF THE MAMMALIAN LUNG AND IS
CC ESSENTIAL FOR NORMAL RESPIRATION.
CC SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC SUBCELLULAR LOCATION: EXTRACELLULAR.
CC MISCELLANEOUS: PULMONARY SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEINS (SP-A AND SP-D) AND 2 SMALL
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-B AND SP-C).
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; M13696; AAA60211.1; -
DR EMBL; K03475; AAA36520.1; -
DR EMBL; M30838; AAA36510.1; -
DR EMBL; M68519; AAA60919.1; -
DR PIR; A24622; LNHUP5.
DR PIR; A25720; LNHUP6.
DR PIR; B25720; LNHUP1.
DR HSP; P19999; 1MSB.
DR MIM; 178630; -
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW Signal; Lactin; Collagen; Repeat; Polymorphism.
FT SIGNAL 1 20
FT CHAIN 21 248 PULMONARY SURFACTANT-ASSOCIATED PROTEIN
FT DOMAIN 28 100 COLLAGEN-LIKE.
FT DOMAIN 153 248 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 155 246 BY SIMILARITY.
FT DISULFID 224 238 BY SIMILARITY.
FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (PROBABLE).
FT VARIANT 9 9 N -> T.
FT VARIANT 66 66 /FTID-VAR_004184.
FT VARIANT 73 73 /FTID-VAR_004185.
FT VARIANT 81 81 /FTID-VAR_004186.
FT VARIANT 85 85 /FTID-VAR_004187.
FT VARIANT 85 85 C -> R.
FT VARIANT 19 19 /FTID-VAR_004188.
FT CONFLICT 19 19 A -> V (IN REF. 2).
FT CONFLICT 45 45 D -> H (IN REF. 2).
```

FT CONFLICT 50 50 L -> V (IN REF. 3).
FT CONFLICT 54 54 P -> L (IN REF. 2).
FT CONFLICT 91 91 P -> A (IN REF. 3).
FT CONFLICT 100 100 P -> R (IN REF. 2).
FT CONFLICT 247 247 E -> D (IN REF. 3).
SQ SEQUENCE 248 AA; 26214 MW; 6A9F0C348BF3633 CRC64;

alignment_scores:
Quality: 63.00 Length: 61
Ratio: 2.250 Gaps: 4
Percent Similarity: 45.902 Percent Identity: 32.787

alignment_block:

US-09-049-696-3 x PSPA_HUMAN

Align seg 1/1 to: PSPA_HUMAN from: 1 to: 248

33 COTCCAGGTAAATGATGAA...CCCTACACTGACGAGATG...GGCAACTG 76
|||||
69 ProProGlyAsnAspGlyLeuProGlyAlaProGlyIleProGlyGluCy 85
|||||
77 TCGAGAGAGGTGAAGAGTCCACTCACCTGCTTCATTCATTCAGGAA 126
|||||
85 sGlyGlyGlyGlyGluPro..... 91

127 AAAAGTTAGCTGAATATGACCAACAGGT..... 155
|||||
92GlyGluArgGlyProProGlyLeuProAlaHisLeuAspGlu 105
156AGGCGATTGTCCATGAGTGGGCTCAT 182
|||||
106 GluLeuGlnAlaThrLeuHisAspHeaPArgHis 116

seq_name: SwissProt_39:PGK_APLCA

seq_documentation_block:

ID PGK_APLCA STANDARD; PRT; 412 AA.
AC O61471;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE PHOSPHOGLYCERATE KINASE (EC 2.7.2.3).
GN PGK.

OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspeidea;
OC Aplysiidae; Aplysia.
OX NCBI_TaxID=6500;
RN [1]

RA Zwartjes R.E., West H., Hattar S., Ren X., Noel F., Nunez-Regueiro M.,
RA Macphree K., Homayouni R., Crow M.T., Byrne J.H., Eskin A.;
RT "Identification of specific mRNAs affected by treatments producing
RT long-term facilitation in Aplysia."
RL Learning Memory 4:478-495(1998).

CC -!- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE = ADP +
CC 3-PHOSPHO-D-GLYCEROYL PHOSPHATE.
CC -!- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AF042738; AAC13267.1;
CC InterPro; IPR001576; PGK.
CC Pfam; PF00162; PGK; 1.

LR PRINTS; PR00477; PHGLYCKINASE.
DR PROSITE; PS00111; PGLYCERATE_KINASE; 1.
KW Transferase; Kinase; Glycolysis.
SQ SEQUENCE 412 AA; 43652 MW; 99B63AA7A2CC2D02 CRC64;

alignment_scores:
Quality: 63.00 Length: 29
Ratio: 3.000 Gaps: 0
Percent Similarity: 72.414 Percent Identity: 44.828

alignment_block:

US-09-049-696-3 x PGK_APLCA

Align seg 1/1 to: PGK_APLCA from: 1 to: 412

81 GAGAAGGTGAAAGGATCCACCTCATCTTCATTTCATTCAGGAAAAA 130
|||||
270 GluLysGlyValLysLeuHisLeuProSerAspHelleThrGlyaspLy 286
|||||
131 GTTAGCTGAATATGACCAACAGGTAGGCGATTGTC 167
|||||
286 sPheAlaaspValAlaLysSerAlaThrAlaSerVal 298

seq_name: SwissProt_39:PGK_OPISI

seq_documentation_block:

ID PGK_OPISI STANDARD; PRT; 415 AA.
AC P50311;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PHOSPHOGLYCERATE KINASE (EC 2.7.2.3).
GN PGK.

OS Opisthorchis sinensis (Clonorchis sinensis).
OC Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
OC Rhabditophora; Eulcithophora; Revertospermata; Mediofusata;
OC Neodermata; Trematoda; Digenea; Opisthorchiida; Opisthorchiata;
OC Opisthorchioidea; Opisthorchiidae; Opisthorchis.

CC NCBI_TaxID=41765;

CC SEQUENCE FROM N.A.

CC Hong S.J.;

CC Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

CC -!- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE = ADP +

CC 3-PHOSPHO-D-GLYCEROYL PHOSPHATE.

CC -!- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).

CC -!- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; L47982; AAD10200.1;

CC HSP; P00560; IQPG.

CC InterPro; IPR001576; PGK.

CC Pfam; PF00162; PGK; 1.

CC PRINTS; PR00477; PHGLYCKINASE.

CC PROSITE; PS00111; PGLYCERATE_KINASE; 1.

CC Transferase; Kinase; Glycolysis.

CC SEQUENCE 415 AA; 44658 MW; 13168A6FB1FB20A0 CRC64;

alignment_scores:

Quality: 63.00 Length: 45
Ratio: 2.423 Gaps: 1
Percent Similarity: 57.778 Percent Identity: 42.222

alignment_block:

US-09-049-696-3 x PGK_OPISI ..

Align seg 1/1 to: PGK_OPISI from: 1 to: 415

```
45 GATGAACCC.....TACACTGACGATGGCAACTGTGGAGA 82
|||||
257 AspGluProGlyAlaAlaIleValAspGluTrpMetGlnLeuAlaLysSe 273
83 GAAGGTGAAGATCCACTCTCTGATTCATTGCGAGGAAAAAGT 132
|||||
273 rLysAsnValLysIleHisLeuProValAspPheIleThrGlyAspLysP 290
133 TACGTGAATATGGACACACAGGTAGGCATTTC 167
|||||
290 heLaGluAspAlaAsnThrGlyThrAlaThrVal 301
```

seq_name: SwissProt_39:CDRL1_CANAL

seq_documentation_block:

```
ID CDRL1_CANAL STANDARD; PRT; 1501 AA.
AC P43071;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MULTIDRUG RESISTANCE PROTEIN CDRL1.
GN CDRL1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 64385 / 1001;
RX MEDLINE=95339406; PubMed=7614555;
RA Prasad R., de Wergifosse P., Balzi E., Goffeau A.;
RT "Molecular cloning and characterization of a novel gene of Candida
albicans, CDRL1, conferring multiple resistance to drugs and
antifungals.";
RL Curr. Genet. 27:320-329(1995).
CC -!- FUNCTION: TRANSPORTER, WHOSE PHYSIOLOGICAL FUNCTION IS NOT YET
ESTABLISHED. CONFERS RESISTANCE TO THE CHEMICAL CYCLOHEXIMIDE.
CC -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
(ABC TRANSPORTERS). PDR5 SUBFAMILY.
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
DR EMBL; X77589; CAA54692.1; .
DR InterPro; IPR003593; AAA.
DR InterPro; IPR003439; ABC_transportr.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00005; ABC_tran; 2.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transmembrane; Glycoprotein; Transport; Repeat.
FT DOMAIN 1 513 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 514 534 POTENTIAL.
FT TRANSMEM 549 569 POTENTIAL.
FT TRANSMEM 598 618 POTENTIAL.
FT TRANSMEM 623 643 POTENTIAL.
FT TRANSMEM 655 675 POTENTIAL.
FT TRANSMEM 765 785 POTENTIAL.
FT DOMAIN 786 1195 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1196 1216 POTENTIAL.
FT TRANSMEM 1230 1250 POTENTIAL.
FT TRANSMEM 1281 1301 POTENTIAL.
FT TRANSMEM 1315 1335 POTENTIAL.
```

```
FT TRANSMEM 1357 1377 POTENTIAL.
FT TRANSMEM 1467 1487 POTENTIAL.
FT NPLBIND 895 902 ATP (POTENTIAL).
SQ SEQUENCE 1501 AA; 169937 MW; BB1440BAD7ED233 CRC64;
```

```
alignment_scores:
Quality: 63.00 Length: 45
Ratio: 2.032 Gaps: 1
Percent Similarity: 68.889 Percent Identity: 33.333
```

alignment_block:

US-09-049-696-3/rev x CDRL1_CANAL

Align seg 1/1 to: CDRL1_CANAL from: 1 to: 1501

```
163 AATGCCCTACCTTGTGTCATATTCAGTAACATTTTCTCTGCAATGAA 114
|||||
482 AsnThrArgProAlaSerProTyrThrValSerPheMetGlnValAr 498
113 ATCAGGAGTGC.....AGTGGATFCCTTTACCCCTTCTCTCCACAGTTGC 70
: ||||| : : : : :
498 gTyrGlyValAlaAlaArgAsnPheLeuArgMetLysGlyAspProSerIleP 515
|||||
69 CCATCTCTCAGTGTAGGGTTCATCATCTACCTGGA 35
|||||
515 rotLePheSerValPheGlyGlnLeuValMetGly 526
```

seq_name: SwissProt_39:Y085_METUA

seq_documentation_block:

```
ID Y085_METUA STANDARD; PRT; 373 AA.
AC Q57550;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ0085.
GN MJ0085.
```

```
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
```

OX NCBI_TaxID=2190;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

RX MEDLINE=96337999; PubMed=8688087;

```
NA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Kl  nk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
```

```
RT Science 273:1058-1073(1996).
```

RL Science 273:1058-1073(1996).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; U67466; AAB98067.1; .

DR TIGR; MJ0085; .

DR InterPro; IPR002491; Peripla_BP.

DR Pfam; PF01497; Peripla_BP_2; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 373 AA; 41812 MW; 8E64F5683E830878 CRC64;

```
alignment_scores:
  Quality: 62.50      Length: 88
  Ratio: 1.359       Gaps: 5
  Percent Similarity: 52.273      Percent Identity: 26.136

alignment_block:
  US-09-049-696-3 x Y085_METJA ..
  Align seg 1/1 to: Y085_METJA from: 1 to: 373

42 AATGATGAACCTACACTGAGGCACTGAGGCAAGGTTGA 91
   :::::  ::  :::::  |||  ||||
208 AspAspLysLysProSerValTyrValGlyGlyTyrLysGly.. 223
   :::::  ::  :::::  |||  ||||
92 AAGGATCCAC.....CTCACTCTGATTTCATTG 120
   :::::  ::  :::::  |||  ||||
224 ....LeuHisGlyIleAspSerThrGluCysLysTyrProProPheValA 239
   :::::  ::  :::::  |||  ||||
121 CA.....GGAAAAAGTTAGCT...GAATATGGACCAAGGTAGGCA 161
   :::::  ::  :::::  |||  ||||
239 laValAsnAlaLysAsnValAlaAspGluLeuGlyLysGluGlyHisVal 255
   :::::  ::  :::::  |||  ||||
162 TTGTGCCATGAGTGGCTCATCTAGATGGGA..... 194
   :::::  ::  :::::  |||  ||||
256 PheValThrLysGluGlnLeuLysTyrAsnProAspIleIlePheIl 272
   :::::  ::  :::::  |||  ||||
195 .....GTATTGACGAGTACATATATATGATGAGA 222
   :::::  ::  :::::  |||  ||||
272 eaSpGluGlyGlyLeuLysLeuValValGluAspTyrLysArgAsnLysG 289
   :::::  ::  :::::  |||  ||||
223 AATTCTACTTATCC 236
   :::::  ::  :::::  |||  ||||
289 luPheTyrAsnSer 293

seq_name: SwissProt_39:YN87_YEAST

seq_documentation_block:
  ID YN87_YEAST STANDARD; PRT; 212 AA.
  AC P53719;
  DT 01-OCT-1996 (Rel. 34, Created)
  DT 01-OCT-1996 (Rel. 34, Last sequence update)
  DT 01-NOV-1997 (Rel. 35, Last annotation update)
  DE HYPOPHYSICAL 23.1 KDA PROTEIN IN URK1-SMM1 INTERGENIC REGION.
  GN YN8714W OR N2060.
  OS Saccharomyces cerevisiae (Baker's yeast).
  OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
  OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
  OX NCBI_TaxID=4932;
  RN [1]
  RP SEQUENCE FROM N.A.
  RA Maurer C.T.C., Urbanus J.H.M., Planta R.J.;
  RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
  RN [2]
  RP SEQUENCE FROM N.A.
  RA Duesterhoeft A., Floeth M., Fritz C., Heuss-Neitzel D.,
  RA Hilbert H., Moestl D.;
  RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
  RN [3]
  RP SEQUENCE OF 1-26 FROM N.A.
  RC STRAIN=S288C / FY1679;
  RX MEDLINE=95208356; PubMed=7900425;
  RA Verhassel P., Aert R., Voet M., Volckaert G.;
  RT "Twelve open reading frames revealed in the 23.6 kb segment flanking
  RT the centromere on the Saccharomyces cerevisiae chromosome XIV right
  RT arm.";
  RL Yeast 10:1355-1361(1994).
  CC -!- SIMILARITY: SOME, TO YEAST YMR206W.
  CC
  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
```

```
modified and this statement is not removed. Usage by and for commercial
entitles requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: 271629; CAA96291.1; -.
DR SGD: S0005297; YNR014W.
KW Hypothetical protein.
FT DOMAIN 30 42 POLY-SER.
SQ SEQUENCE 212 AA; 23076 MW; C92421056E7BFE7C CRC64;

alignment_scores:
  Quality: 62.00      Length: 54
  Ratio: 2.000       Gaps: 1
  Percent Similarity: 57.407      Percent Identity: 33.333

alignment_block:
  US-09-049-696-3/rev x YN87_YEAST ..
  Align seg 1/1 to: YN87_YEAST from: 1 to: 212

195 CTCGCCATCGTAGATGAGCCACATCATGACAAATGCCCTACCTGTGGT 146
   |||||  |||||  ::  :::::  |||
140 LeuProSerPheAspAsnArgSerSerSerGluMetLysArgSerVal 156
   :::::  ::  :::::  |||  ||||
145 CCATATTTCAG.....CTAACTTTTTCCTGCAATGAAATCAGAGTGAG 102
   :::::  ::  :::::  |||  ||||
156 rTyAlaGlnHisSerMetMetPheProIleSerAspGlnGluProG 173
   :::::  ::  :::::  |||  ||||
101 GTGGATCCTTTTACCCCTTCTCTCCACAGTTGCCCATCTGCTCACTAGG 52
   |||  :::::  |||||  |||||  :::::  |||
173 lnThrSerAlaSerProAsnAspHisSerAspProSerCysProCysAsn 189
   :::::  ::  :::::  |||  ||||
51 GTTCATCATTTAC 40
   :::::  ::  :::::  |||  ||||
190 ArgHisHisHis 193

seq_name: SwissProt_39:MCAS_MYCBO

seq_documentation_block:
  ID MCAS_MYCBO STANDARD; PRT; 2110 AA.
  AC Q02251;
  DT 01-JUN-1994 (Rel. 29, Created)
  DT 01-JUN-1994 (Rel. 29, Last sequence update)
  DT 15-JUL-1998 (Rel. 36, Last annotation update)
  DE MYCOCEROSIC ACID SYNTHASE.
  GN MAS.
  OS Mycobacterium bovis.
  OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
  OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
  OX NCBI_TaxID=1765;
  RN [1]
  RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
  RX MEDLINE=92406887; PubMed=1527058;
  RA Mathur M., Kolattukudy P.E.;
  RT "Molecular cloning and sequencing of the gene for mycocerosic acid
  RT synthase, a novel fatty acid elongating multifunctional enzyme, from
  RT Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin.";
  RL J. Biol. Chem. 267:19388-19395(1992).
  CC -!- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
  CC : METHYLMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
  CC : FORM MYCOCEROSYL LIPIDS.
  CC -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE.
  CC -!- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
  CC : ARRANGEMENT.
  CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
  CC -!- SIMILARITY: PARTIAL TO S. ERYTHRAEA ERYTHRONOLIDE SYNTHASE, MODULE
  CC : 4, AND TO VERTEBRATE FATTY ACID SYNTHASES.
  CC
  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
```


modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

```
CC -----
CC EMBL; M95808; AAA25369.1;
CC PIR; B44110; B44110.
CC HSP; Q02054; IAF8.
CC InterPro; IPR001227; Acyltransf_domain.
CC InterPro; IPR002085; Adh_zinc.
CC InterPro; IPR000794; ketoacyl-synt.
CC InterPro; IPR003880; Phosphopant_attach.
CC Pfam; PF00698; Acyl_transf; 1.
CC Pfam; PF00107; adh_zinc; 1.
CC Pfam; PF00109; ketoacyl-synt; 1.
CC Pfam; PF00550; pp-binding; 1.
CC PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
CC PROSITE; PS00606; B-KETOACYL SYNTHASE; 1.
CC PROSITE; PS00075; ACP_DOMAIN; 1.
CC Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
KW Transferase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.
FT DOMAIN 1 430 BETA-KETOACYL SYNTHASE.
FT DOMAIN 533 852 ACYL TRANSFERASE.
FT DOMAIN ? ? ENOYL REDUCTASE.
FT DOMAIN ? ? BETA-KETOACYL REDUCTASE.
FT DOMAIN 2026 2096 ACYL CARRIER (ACP).
FT ACT_SITE 177 177 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT ACT_SITE 623 623 ACYL TRANSFERASES (BY SIMILARITY).
FT NP_BIND 1561 1578 NADP (ER).
FT NP_BIND 1765 1780 NADP (KR).
FT BINDING 2059 2059 PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ SEQUENCE 2110 AA; 225577 MW; CC658215D7155300 CRC64;

alignment_scores:
  Quality: 62.00 Length: 58
  Ratio: 2.067 Gaps: 3
Percent Similarity: 51.724 Percent Identity: 34.483

alignment_block:
US-09-049-696-3 x MCAS_MYCBO
Align seg 1/1 to: MCAS_MYCBO from: 1 to: 2110

78 GGAGAGAGGGTGAAGGATCCACCTC.....AC 106
||||| ||||| :||| :||| :||| :||| :|||
1205 GlycylalanylSerTrpLeuLeuIleAspThrSerAsnAlaValAspTh 1221
||||| :||| :||| :||| :||| :||| :|||
107 TCCTGATTTTCATGTCAGGA.....AAAAAGTTAGCTGAATATGGAC 147
||||| :||| :||| :||| :||| :||| :|||
1221 rProAspMetLeuAlaSerThrLeuThrAspAlaLeuLysSerHisGlyP 1238
||||| :||| :||| :||| :||| :||| :|||
148 CACAAGGTAGGCGATTGTCATGATGGCTCATCTACGATGGGAGTA 197
||||| :||| :||| :||| :||| :||| :|||
1238 roGlnGlyThr.....GlucyAlaSerLeuSerTrpSerVal 1250

198 TTTCAGCAGTACATAATGATGAG 221
||| ||||| :||| :||| :||| :||| :|||
1251 GlnAspThrProAsnAspGln 1258

seq_name: SwissProt_39:ATH1_YEAST
seq_documentation_block:
ID ATH1_YEAST STANDARD; PRT; 1211 AA.
DC P48016;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE VACUOLAR ACID TREHALASE PRECURSOR (EC 3.2.1.28) (ALPHA, ALPHA-
DE TREHALASE) (ALPHA, ALPHA-TREHALOSE GLUCOHYDROLASE).
GN ATH1 OR YPR026W OR YP9367.06.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
```

```
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 130-1211 FROM N.A.
RC STRAIN=MM303-1B;
RX MEDLINE=96076626; PubMed=7502577;
RA DeStruelle M., Holzer H., Klionsky D.J.;
RT "Isolation and characterization of a novel yeast gene, ATH1, that is
RL required for vacuolar acid trehalase activity.";
RN Yeast 11:1015-1025(1995).
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Edcock K., Churcher C.M., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ALPHA,ALPHA-TREHALOSE + H(2)O = 2 D-GLUCOSE.
CC -!- SUBCELLULAR LOCATION: LYOSOME-LIKE VACUOLES (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO FAMILY 65 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
```

```
EMBL; X84156; CAA58961.1;
DR EMBL; 249274; CAA89280.1;
DR EMBL; 271255; CAA95022.1;
DR SGD; S0006230; ATH1.
KW Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1
FT CHAIN ? 1211
FT CARBOHYD 28 28 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 32 32 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 207 207 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 247 247 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 255 255 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 325 325 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 370 370 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 376 376 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 488 488 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 539 539 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 628 628 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 696 696 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 705 705 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 879 879 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 910 910 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 972 972 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1031 1031 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1064 1064 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1147 1147 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1157 1157 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1168 1211 LOVGDKGNTDKRKTIVVAVGVYDDDDKNGKATIKETV
FT LND -> FAGG (IN REF. 1).
SQ SEQUENCE 1211 AA; 136920 MW; 34AEC44B0B648DEC CRC64;

alignment_scores:
  Quality: 61.50 Length: 81
  Ratio: 1.500 Gaps: 4
Percent Similarity: 50.617 Percent Identity: 24.691

alignment_block:
US-09-049-696-3 x ATH1_YEAST
```

```
Align seg 1/1 to: ATH1_YEAST from: 1 to: 1211
12 GTTCTGGTGTGAGTCTACTCTCCAGTAATGATGAACCTTACACTGA 61
1127 VallylleleuAspAsnValProValSerProSerGluProTyrTyrG1 1143
62 GCAGATGGCGCAACTGTGCGAGAGAAGGTGAAGAGTCCACTCACCTCTG 111
1143 uAlaserArgasn.....GlnSerLeuIleGluIleLeuProS 1156
112 AT.....TTTCATTCAGGAAAAAGTTAGCTGAATATGGACCA 149
1156 eAsnArThrThrPheThrIleAspTyrAspLysLeuGlnValGlyAsp 1172
150 CAAGTAGGCGATTGTCCATGATGGCTCATCTAGCATGG.....191
1173 LysGlyAsnThr.....AspTyrArgLysThrArgTyrIleValVa 1186
192 .....CGAGTATTGACGAGTACAATAATGATGAGAAA 224
1186 lAlaValGlnGlyValTyrAspTyrAspAspAsnLys 1200
seq_name: SwissProt_39:ZEP1_MOUSE
seq_documentation_block:
ID ZEP1_MOUSE STANDARD; PRT; 2688 AA.
AC Q03172;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE ZINC FINGER PROTEIN 40 (TRANSCRIPTION FACTOR ALPHAA-CRYBP1) (ALPHA A-
DE CRYSTALLIN-BINDING PROTEIN I) (ALPHA A-CRYBP1).
GN HIVEP1 OR ZNF40 OR CRYABP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=DBA/2J;
RC MEDLINE=95138112; PubMed=7836383;
RA Nakamura T., Donovan D.M., Hamada K., Sax C.M., Norman B.,
RA Brady J.P., Kantorow M., Sax C.M., Westphal H., Platiogorsky J.;
RT "Murine transcription factor alpha A-crystallin binding protein. I.
RT Complete sequence, gene structure, expression, and functional
RT inhibition via antisense RNA."
RL J. Biol. Chem. 270:1221-1229(1995).
RN [2]
RP SEQUENCE OF 2024-2688 FROM N.A.
RC TISSUE=Lens epithelium;
RX MEDLINE=90287161; PubMed=1694016;
RA Flanagan J.R., Ozato K., Westphal H., Platiogorsky J.;
RA "Regulation of the mouse alpha A-crystallin gene: isolation of a cDNA
RT encoding a protein that binds to a cis sequence motif shared with the
RT major histocompatibility complex class I gene and other genes."
RL Mol. cell. Biol. 10:3700-3708(1990).
CC -1- FUNCTION: TRANSCRIPTION FACTOR WHICH BINDS SPECIFICALLY TO THE
CC -1- PALINDROMIC SEQUENCE 5'-GGGAAATCCC-3' IN THE ALPHA-A CRYSTALLIN
CC PROMOTER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY
CC SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH
CC ZINC-FINGER IN-BETWEEN.
CC -1- SIMILARITY: STRONG, TO HIVEP2.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

Score	Strd	Orig	ZScore	EScore	Len	Documentation
Sequence						
pir2:JG0168	+	356.00	751.65	7.8e-35	913	! gob-5 protein - mouse
pir2:TJ20205	+	255.00	531.52	1.4e-22	905	! Lu-ECAM-1 protein - bovine
pir2:DB3347	+	73.50	140.12	1.50	544	! hypothetical protein PA2392 [im
pir2:T02319	+	73.00	140.27	1.71	470	! hypothetical protein AT2g34310
pir2:S44283	+	71.50	145.46	2.38	173	! pail repressor homolog - Lactob
pir2:C70612	+	68.50	125.30	6.46	1034	! probable recB protein - Mycobac
pir2:T39489	+	67.50	128.67	7.91	449	! L-sorbose dehydrogenase - Ace
pir2:A96719	+	67.50	121.65	8.50	1028	! hypothetical protein T6C23.9 [im
pir2:AS3962	+	67.00	126.58	9.19	505	! endoglucanase - Erwinia carotov
pir2:T35845	+	66.50	124.60	10.66	561	! hypothetical protein SC9A10.09
pir2:T18650	+	66.50	121.22	11.04	836	! hypothetical protein B0035.12 -
pir2:C31915	+	66.00	128.67	12.88	993	! lantibiotic nisin biosynthesiz
pir2:T06384	-	65.00	124.43	15.72	389	! probable ethanolaminephosphotr
pir2:JC6093	-	64.50	116.23	19.44	901	! dead ringer nuclear protein - f
pir2:DB3320	+	64.00	123.72	20.48	327	! hypothetical protein PA2809 [im
pir2:F81688	-	63.50	128.79	22.12	158	! hypothetical protein TC0573 [im
pir2:F94726	-	63.50	115.77	25.26	735	! probable Na+/H+ antiporter [imp
pir1:LNHUP5	+	63.00	123.88	26.45	248	! pulmonary surfactant protein A
pir1:LNHUP6	+	63.00	123.88	26.45	248	! pulmonary surfactant protein A
pir1:JC4936	+	63.00	119.71	27.60	406	! biphenyl dioxygenase [EC 1.14. -
pir2:S57198	-	63.00	108.63	30.89	1501	! multidrug resistance protein - C
pir2:F72423	-	62.50	119.54	31.44	364	! hypothetical protein TM0053 - F
pir2:E64310	+	62.50	119.33	31.51	373	! iron transport system binding p
pir2:S63342	-	62.00	123.03	34.51	212	! hypothetical protein YNR014W -
pir2:C71250	+	62.00	111.87	38.67	792	! probable cation-transporting AT
pir2:B44110	+	62.00	103.57	42.08	2111	! mycoserolate synthase [EC 2.3. -
pir2:A70668	+	62.00	103.56	42.08	2111	! mycoserolate synthase (EC 2.3. -
pir2:T27362	-	61.50	120.02	40.47	266	! hypothetical protein Y71A12C.2
pir2:H72529	+	61.50	119.14	40.84	295	! hypothetical protein APE2212 -
pir2:S54500	+	61.50	107.18	46.13	1211	! alpha,alpha-trehalase [EC 3.2.2 -
pir2:T49477	+	61.50	100.42	49.42	2688	! alpha-A-crystallin-binding pro
pir2:F31201	+	61.00	120.42	45.84	223	! GII-related finger protein HKR4
pir2:S32427	+	61.00	118.83	46.59	269	! triose-phosphate isomerase [EC
pir2:G95619	+	60.50	117.93	53.48	263	! probable transport system perme
pir2:DB3124	+	60.50	111.40	57.16	569	! hypothetical protein PA4163 [im
pir2:A47704	+	60.50	107.71	59.35	879	! endoglucanase I [EC 3.2.1.-) Cel
pir2:DB4827	+	60.00	117.69	60.98	238	! probable protein kinase [import
pir2:T22770	+	60.00	116.59	61.66	271	! hypothetical protein F56C4.1 -
pir2:A38415	+	60.00	112.28	64.43	451	! macrophage scavenger receptor,
pir2:E92340	+	60.00	112.15	64.52	458	! probable glucanase permealase VCC
pir2:T35928	-	60.00	111.96	64.64	468	! probable aldehyde dehydrogenase


```

alignment_block: ..
US-09-049-696-3 x T18650
Align $seg 1/1 to: T18650 from: 1 to: 836

```

72 AACTGTGGAGAGAGGTTGAAAGGATCCACTCCTCATCTCGATTTCATTC 121
73
349 AsnSerAlaGluGluGlyArgSerLeuArgThrTyrAlaPheLeuLe 365
122 AGAAAAAAGCTTA.....CCTCAATATGACCA..... 149
123
365 uArgArgIleHisLeuThrGlySerSerAspTyrSerPrometAlaG 382
150CAAGGTAGGGCATTTGTCCATGAGTGGCTCATCTACGA 188

189 TGGGGAGTATTTGACGAGTACANTATGTATGAGAATTCTACTATTCCAA 238
||| ||| :|||||::: ||| :
306 luvaipnenspgugugyaataaleudeuarigouiripnesermetatba 399

239 T 239 :

413 F 413

```
seq_name: p112:G31913
seq_documentation_block:
  antibiotic_nisin_biosynthesis_protein_nisB - Lactococcus lactis
```

N;Alternate names: probable membrane-associated protein A
C;Species: *Lactococcus lactis*
C;Date: '21-May-1990 #sequence_revision 13-Aug-1999 #text_change
C;Accession: S36735.C31915.C48951.C45821.B43743.S36985

R.; Kuipers, O.P.; Beerthuyzen, M.M.; Siezen, R.J.; de Vos, W.M.
Eur. J. Biochem. 216, 281-291, 1993
A:Title: Characterization of the nisin gene cluster nisABTCIPR
A:Reference number: 536734, MMRN-93272027

A;Accession: S36735
A;Status: preliminary
A;Molecule type: DNA

A: Cross-references: GB:L16226; NID:g400365
R: Buchman, G.W.; Banerjee, S.; Hansen, J.N.

J. Biol. Chem. 263, 16260-16266, 1988

A; Accession: C31915

A: Residues: 1-63 <BUC>

A: Cross-references: GB: J04057; NID: q153816; PIDN: AAA88607.1; PIDN: RESIDUES: 1-93 <DOC>

R;Engelke, G.; Gutowski-Eckel, Z.; Hammelmann, M.; Entian, K.D.

Appl. Environ. Microbiol. 58, 3730-3743, 1992

A;Title: Biosynthesis of the lantibiotic nisin: genomic organization

A;Reference number: A48951; MUID:93128945

A;Accession: C48951

A;Status: preliminary

A; Molecule type: DNA

A; Residues: 1-18, 'C', 20-655, 'K', 657-894, 'T', 896-993 <ENG>

A;Cross:references: GB:X68307; GB:S52234; NID:g44040; PIDN:CAA48

λ ; Experimental source: strain 6F3

A; Note: sequence extracted from NCBI backbone (NCBIN:122292, NCBI

R; Dodd, H.M.; Horn, N.; Gasson, M.J.

J. Gen.: Microbiol. 136, 555-566, 1990

A;Title: Analysis of the genetic determinant for production of

A;Reference number: A45821; MUID:90362041

A;Accession: C45821
4;Status: preliminary

A; status: preliminary
2-molecule type: DNA

A; MOLECULE TYPE: DNA
A: Residues: 1-9 <DON>

A;Residues: I-9 <DOD>
A:Cross-references: CB.M79445

A: CROSS-REFERENCES: GB:M/9443
R: Steen, M. T.: China. Y. J.: Hansen, J. N.

K/Steel, M.I.; Chung, I.J.; Hansen, J.N.	Appl Environ Microbiol	57	1181-1188	1991
--	------------------------	----	-----------	------

Appl.: ENVIRON. MICROBIOL. 37, 1181-1188, 1981

A,11111: Characterization of the *urim* gene as part of a polycistronic

seq_name: pir2:D83320

seq_documentation_block:
hypothetical protein PA2609 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83320
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: D83320
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-327 <STO>
A:Cross-references: GB:AE004689; GB:AE004091; NID:g9948665; PIDN:AAG05997.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2609

alignment_scores:
Quality: 64.00 Length: 69
Ratio: 1.778 Gaps: 3
Percent Similarity: 52.174 Percent Identity: 30.435

alignment_block:

US-09-049-696-3 x D83320 ..

Align seg 1/1 to: D83320 from: 1 to: 327

```
3  ATGCTGATGTTCTGTTGCTGACTCTACTCTCCAGGTAATGATGAACC 52
|||||  :::::  |||
197  AsnAlaAspCysIleLeuGlnSerIlePheHisProGlyTyrGlnGluTh 213
53  CTACACTGACGACATGGCAACTGTGGAGAGAG..... 86
:::  |||  |||:::
213  rHisArgGluAlaSerArgLeuLeuGlyAspArgAlaIleValIleLysG 230
87  ....GGTGAAGGATCCACCTCCTGATTTCATTGCAGGAAAAAAG 131
|||  ::|  ::|  ::|  ::|  ::|  ::|  ::|  ::|  ::|  ::|
230  lyGluGlyGlyGluIleGluValAsnProAspValValAlaHisLeu... 245
132  TTACCTGAATATGCA...CCACAAGTAGGCGATTTCATCAGTGGGC 178
|||||  |||  ::|  ::|  ::|  ::|  ::|  ::|  ::|  ::|
246  .....TyrGlyCysAspGlnGlyGluAlaTrpAspGluGluTrpPr 259
179  TCATCTA 185
|||
259  oAlaLeu 261
```

This Page Blank (uspio)

OM of: US-09-049-696-3 to: Issued_Patents_AA.* out_format : pfs

Date: Mar 30, 2002 2:32 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet_n2p.model -DEV=rlp
-Q=/cgn2_1/USPTO/US09049696/runat_28032002_145238_2012/app_query.fasta_1.12579
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum2
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696_cgn1_109 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-3
Query length: 240
Database: Issued_Patents_AA.*
Database sequences: 212252
Database length: 22503292
Search time (sec): 362.110000

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/iaa/PCRTUS_COMB.ppt:US93-02475-12		61.00	130.85	4.54	171
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-453-117-2		60.00	118.38	8.52	451
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-948-222-2		60.00	118.38	8.52	451
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-973-145-2		60.00	118.38	8.52	451
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-276-400-10		60.00	118.38	8.52	451
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-448-076-10		60.00	118.38	8.52	451
/cgn2_6/ptodata/2/iaa/PCRTUS_COMB.ppt:US93-02475-12		60.00	118.38	8.52	451
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-09-036-987A-6		60.00	92.28	19.55	5588
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-370-700-6		59.50	92.28	19.55	5588
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-09-314-242-2		59.00	113.92	11.04	617
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-022-835-4		59.00	109.30	14.37	857
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-388-809-4		59.00	109.30	14.37	857
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-647-714-4		59.00	109.30	14.37	857
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-479-722B-2		59.00	101.42	18.46	1833
/cgn2_6/ptodata/2/iaa/PCRTUS_COMB.ppt:US93-02475-12		59.00	101.42	18.46	1833
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-965-056-15		58.00	118.97	13.34	267
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-09-034-177-1		58.00	110.89	17.25	582
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-022-835-2		58.00	106.89	19.59	856
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-388-809-2		58.00	106.89	19.59	856
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-375-100-1		58.00	106.89	19.59	856
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-647-714-2		58.00	106.89	19.59	856
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-07-956-483-11		58.00	106.89	19.59	856
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-468-557-2		58.00	105.35	20.58	993
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-872-961A-4		56.50	115.18	21.37	271
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-09-231-258-4		56.50	115.18	21.37	271
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-07-992-827D-1		56.50	113.36	22.64	323
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-216-593-1		56.50	113.36	22.64	323
/cgn2_6/ptodata/2/iaa/PCRTUS_COMB.ppt:US93-12380-1		56.50	113.36	22.64	323
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-391-339-18		56.50	110.37	24.90	431
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-484-274A-18		56.50	110.37	24.90	431
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-484-493-11		56.50	108.63	26.32	510
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-484-494-11		56.50	108.63	26.32	510
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-212-11		56.50	108.63	26.32	510
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-249-003-11		56.50	108.63	26.32	510
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-462-481-4		56.50	105.66	28.92	679
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-436-771-6		56.50	105.66	28.92	679
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-434-998-6		56.50	105.66	28.92	679
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-487-797-6		56.50	105.66	28.92	679
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-701-005A-4		56.50	105.66	28.92	679
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-479-895-4		56.50	105.66	28.92	679

/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-08-943-956A-4 + 56.50 105.66 28.92 679
/cgn2_6/ptodata/2/iaa/PCRTUS_COMB.ppt:US93-02058-6 + 56.50 105.66 28.92 679
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-08-751-189-4 + 56.50 91.62 45.20 2629
/cgn2_6/ptodata/2/iaa/5B_COMB.ppt:US-09-060-836-4 + 56.50 91.62 45.20 2629
/cgn2_6/ptodata/2/iaa/6B_COMB.ppt:US-09-184-445-4 + 56.50 91.62 45.20 2629

seq_name: /cgn2_6/ptodata/2/iaa/PCRTUS_COMB.ppt:US93-02475-12

seq_documentation_block:
: Sequence 12, Application PC/TUS9302475
: GENERAL INFORMATION:
: APPLICANT: Wisniewski, Bernadine J.
: TITLE OF INVENTION: Tumor Necrosis Factor with Modified
: TITLE OF INVENTION: Ion Channel
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Donald G. Lewis
: STREET: 8328 Regents Road #1E
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.5 inch, 1.44 M storage
: COMPUTER: VE System 386
: OPERATING SYSTEM: MS-DOS 5
: SOFTWARE: Word Perfect
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/02475
: FILING DATE: 19930412
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/852,625
: FILING DATE: 12 March 1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Donald G. Lewis
: REGISTRATION NUMBER: 28636
: REFERENCE/DOCKET NUMBER: BJW-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 554-2421
: TELEFAX: (619) 554-6312
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 171 amino acids
: TYPE: AMINO ACIDS
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FEATURE:
: NAME/KEY: Lymphotoxin (rabbit)
: OTHER INFORMATION: Two blank residues designated by
: OTHER INFORMATION: "aaa" are inserted after residue No. 34 and No. 61
: OTHER INFORMATION: of murine lymphotoxin and the sequence numbering is
: OTHER INFORMATION: augmented by 1 starting with residue No. 35 and
: OTHER INFORMATION: again augmented by 1 starting with residue No. 62
: OTHER INFORMATION: in order to maximize the sequence homology with
: OTHER INFORMATION: human lymphotoxin.
: PUBLICATION INFORMATION:
: AUTHORS: Ito, H., Shirai, T., Yamamoto, S.,
: AUTHORS: Akira, M., Kawahara, S., Todd, C.W.
: AUTHORS: and Wallace, R.B.
: TITLE: Molecular Cloning of the Gene Encoding
: TITLE: Rabbit Tumor Necrosis Factor.
: JOURNAL: DNA
: VOLUME: 5
: PAGES: 157-165
: DATE: 1986
: RELEVANT RESIDUES IN SEQ ID NO: 12: 1-171 (includes
: RELEVANT RESIDUES IN SEQ ID NO: 2 blanks)
PCT-US93-02475-12

alignment_scores: 61.00 Length: 83
Quality:

US-09-448-076-10

alignment_scores:		
Quality:	60.00	Length: 75
Ratio:	1.395	Gaps: 7
Percent Similarity:	57.333	Percent Identity: 33.333

```
alignment_block:      US-09-049-696-3  x  PCT-US96-08081-2
```

Align seq 1/1 to: PCT-US96-08081-2 from: 1 to: 451

35 CCAGGTAAATGATGAACCCCTACACTGACGACATGGCAACATTGTGGAGAGAA 85
|||||::: |||
326 ProglyAsnSerGlyPro.....LysGlyGlnLysGlyGluLy 338
|||::: |||
86 GGGTGAAGAGTCACCACTCCTGCTGATTTTCATTCAGCGAAAAAAGCTTAG 135
|||||::: ||||||||| ||| ::| ::| ::|
338 sGlySerClyAsnThrLeuThrPro...PheThrLysValArgLeuValG 354
|||::: |||::: |||::: |||::: |||::: |||:::
136 CTGAATATGGACCA...CAAGGTAGGCCA...TTTGTCCTAT.....GAG 173
|||::: |||::: |||::: |||::: |||::: |||:::
354 LyGlySerGlyProHisGluGlyArgValGluIleLeuHisSerGlyGln 370
|||::: |||::: |||::: |||::: |||::: |||:::
174 TGGCTCATCTA.....CGATGGGAGCTATTTCACGACTACAATAA 214
|||||::: |||::: |||::: |||::: |||::: |||:::
371 TrpGlyThrIleCysAspAspArgTrp.GluValArgValGlyGlnValy 387
|||::: |||::: |||::: |||::: |||::: |||:::
215 TGATGAGAAAATTCCTACTTATCCA 237
|||::: |||::: |||::: |||::: |||::: |||:::
387 alCysArgSerLeuGlyThrPro 394

seq_name: /cqn2_6/ptodata/2/iaa/6B_COMB.pcp:US-09-036-987A-6

```
seq_documentation_block:
  : Sequence 6, Application US/09036987A
  : Patent No. 6143526
  :
  : GENERAL INFORMATION:
  :
  : APPLICANT: Baltz, Richard H.
  : APPLICANT: Broughlon, Mary C.
  : APPLICANT: Crawford, Kathryn P.
  : APPLICANT: Madduri, Krishnamurthi
  : APPLICANT: Merlo, Donald J.
```

```

** APPLICANT: Madduri, Krishnamurthy
** APPLICANT: Treadway, Patti J
** APPLICANT: Turner, Jan R
** APPLICANT: Walldron, Clive
** TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
** FILING REFERENCE: 50489 DIV1
** CURRENT APPLICATION NUMBER: US/09/370,700
** CURRENT FILING DATE: 1999-08-09
** EARLIER APPLICATION NUMBER: US 09/36987
** EARLIER FILING DATE: 1998-03-09
** NUMBER OF SEQ ID NOS: 39
** SOFTWARE: patentin ver. 2.0
** SEQ ID NO 6
** LENGTH: 5388
** TYPE: PRT
** ORGANISM: Saccharopolyspora spinosa
** US-09/370-700-6

```

alignment_scores:			
Quality:	60.00	Length:	59
Ratio:	1.714	Gaps:	3
Percent Similarity:	59.322	Percent Identity:	32.203

```
alignment_block: .
US-09-049-696-3 x US-09-370-700-6
..
Align_seg 1/1 to: US-09-370-700-6 from: 1 to: 5588
```

Align seg 1/1 to: US-09-370-700-6 from: 1 to: 5588

15 CTGGTGTGCTACTACTCT...CCAGGTAATGATGAACCTACACTGA 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2174 LeuLeuThrGLuAsnThrProPrpPGIvAsnSerHisProArqVa 2190

· 15 CTGGTTGCTGAGTCTACTCCT...CCAGGTAATGATGAACCCCTACACTGA 61
 · |||:||||:||||:||||| |||||:||||:||||:||||:
 2174 LeuLeuThrGluAsnThrProTrpProGlyAsnSerHisProArgArgVa 2190

2174 LeuLeuThrGluAsnThrProTrpProGlyAsnSerHisProArqArqVa 2190

62 GCAGATGGGCACCTGTGGAGAGAAGGGTGAAGGATCCACCTC..... 104

[illegible]

2196 |G|vvalserserphec|v|leserg|v|ThrAsnAlaHisvalTleIeng 2207

141 TATGGACCAAGGTTAGGCATTTGTC 167
105ACTCTGATTTTCAGGAAAAAGTTAGCTGAA 141
106TCTGATTTTCAGGAAAAAGTTAGCTGAA 141
2207 LuGlnAlaSerLysThrProAspGluThrAlaAspLysSer..... 2220
2211 ...GlyProAspSerGluSerThrVal 2228

seq_name: /cqn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-314-242-2

```
seq_documentation_block:
; Sequence 2, Application US/09314242A
; Patent No. 6248575
; GENERAL INFORMATION:
```

2. APPLICANT: ELIZABETH J. GOLIGHTLY
 3. TITLE OF INVENTION: NUCLEIC ACIDS ENCODING POLYPEPTIDES
 4. TITLE OF INVENTION: HAVING L-AMINO ACID OXIDASE ACTIVITY
 5. PRIORITY REFERENCE: 5556 200-US

: CURRENT APPLICATION NUMBER: US/09/314,242A
 :
 : CURRENT FILING DATE: 1999-05-18
 :
 : EARLIER APPLICATION NUMBER: 09/080,428
 :
 : EARLIER FILING DATE: 1998-05-18

```

; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 617

```

```

; TYPE: PRT
; ORGANISM: Trichoderma harzianum

```

US-09-314-242-2

alignment scores:

Quality:	59.50	Length:	36
----------	-------	---------	----

Ratio: 2.587

Percent Similarity: 63.889 Percent Identity: 41.667

```

alignment_block:
US-09-049-696-3 x US-08-022-835-4
Align seg 1/1 to: US-08-022-835-4 from: 1 to: 857

16 TGGTTGCTGAGTCTACTCTCCAGGTAATGATGAACCTACACTGAGCAG 65
   |||   :: |||   ::::: |||   |||   |||   ::
615 rTpSerAsnLysSerLeuAspAspIleTrpAsnAsnMetThrTrpMetGl 631
   ::::: |||   ::::: |||   ::::: |||   ::::: |||   ::
66 ATGGGCAACTGTGGCAGAGGAGGTGAAAGGATCCACCTCACTCTCTGATT 115
   |||   |||   |||   ::::: |||   |||   |||   ::
631 nTrp.....GluArgGluIleAspAsnTyThrSerLeuIleTyrs 645
   |||   |||   |||   |||   |||   |||   |||   |||
116 CATTCCAGGAAAAAGTTA.....GCTGAATATGGACCAAGGTAGG 158
   |||   |||   |||   |||   |||   |||   |||   |||
645 erLeuLeuGluLysSerGlnThrGlnGlnGluMetAsnGluGlnGluLeu 661
   |||   |||   |||   |||   |||   |||   |||   |||
159 GCATTGTGTCATGAGTGGGCTCATCTACGATGGGGAGTATTGAGGAGTA 208
   ::::: |||   |||   |||   |||   |||   |||   |||
662 LeuGluLeuAspLysTrpAlaSerLeu...TrpAsnTrpPheAspIleTh 677
   |||   |||   |||   |||   |||   |||   |||
209 CAAT 212
   |||
677 rAsn 678

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-388-809-4

seq_documentation_block:
; Sequence 4, Application US/08388809
; Patent No. 5576000
;
; GENERAL INFORMATION:
; APPLICANT: REITZ, JR., MARVIN S.; FRANCHINI,
; APPLICANT: GENOVEFFA; MARKHAM, PHILIP D. GALLO, ROBERT
; APPLICANT: C.; LORI, FRANCO C.; POPOVIC, MIKULAS; AND
; APPLICANT: GARTNER, SUZANNE
; TITLE OF INVENTION: MOLECULAR CLONES OF HIV-1
; TITLE OF INVENTION: AND USES THEREOF
;
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK, 3.5"
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/388,809
; FILING DATE: 15-FEB-1995
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/022,835
; FILING DATE: 25-FEB-1993
;
; APPLICATION NUMBER: US 07/599,491
; FILING DATE: 17-OCT-1990
;
; ATTORNEY/AGENT INFORMATION:
; NAME: LESLIE A. SERUNIAN
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4092US2
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
;
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 857 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

alignment_scores:	
Quality:	59.00
Ratio:	1.639
Percent Similarity:	52.941
Percent Identity:	35.294
Gaps:	3
Length:	68

? FILING DATE: ?
 ? PRIOR APPLICATION DATA: ?
 ? APPLICATION NUMBER: US/08/388,809 ?
 ? FILING DATE: 15-FEB-1995 ?
 ? APPLICATION NUMBER: US 08/022,835 ?
 ? FILING DATE: 25-FEB-1993 ?
 ? APPLICATION NUMBER: US 07/599,491 ?
 ? FILING DATE: 17-OCT-1990 ?
 ? ATTORNEY/AGENT INFORMATION: ?
 ? NAME: LESLIE A. SERUNIAN ?
 ? REGISTRATION NUMBER: 35,353 ?
 ? REFERENCE/DOCKET NUMBER: 2026-4092US2 ?
 ? TELECOMMUNICATION INFORMATION: ?
 ?

```

: Patent No. 6074840
:
: GENERAL INFORMATION:
:
: APPLICANT: Bonadio, Jeffrey
: APPLICANT: Yin, Wushan
:
: TITLE OF INVENTION: LATENT TGF( BINDING PROTEIN (LTBP)
:
: TITLE OF INVENTION: GENES, COMPOSITIONS AND METHODS
:
: NUMBER OF SEQUENCES: 13
:
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Williams, Morgan & Amerson
: STREET: 7576 Hillmont, Suite 250
: CITY: Houston
: STATE: Texas
: COUNTRY: USA
: ZIP: 77040
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/479,722B
: FILING DATE: 07-JUN-1995
: CLASSIFICATION:
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US95/02251
: FILING DATE: 21-FEB-1995
: PRIOR APPLICATION DATA:

```

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/316,650
 FILING DATE: 30-SEP-1994
 CLASSIFICATION:
 APPLICATION NUMBER: US 08/199,780
 FILING DATE: 18-FEB-1994
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: PARKER, DAVID L.
 REGISTRATION NUMBER: 32,165
 REFERENCE/DOCKET NUMBER: UMIC009P--
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-3000
 TELEFAX: (713) 789-2679
 TELEX: 79-0924
 INFORMATION FOR SEQ ID NO: 18:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1833 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 UNRS-02251-18

alignment_scores:		
Quality:	59.00	Length: 53
Ratio:	2.107	Gaps: 3
Percent Similarity:	52.830	Percent Identity: 33.962

Align seq 1/1 to: US-08-479-722B-2 from: 1 to: 1833

33 CCTCCAGGTAATGATGAACCTTACACTGAGCAGATGGGCAACTGT..... 77
||||| :::::||| ||::: ||::: ||
1637 proProArdSerSerGluValTyrAla...GlnLeuCysAsnValAlaAr 1652

79GGAGAGAGGCTCAAGGATCCACCTCCTCGATTTCATG 120
:::|||||::| |||||::| |::| ::
1652 glieGluaGluArgGlyAlaGlyIleHisPheArgProGlyTyrGluT 1669

1686 Alaprophe 1688

seq_documentation_block:

```

; sequence 18, Application PC/1039502251
; GENERAL INFORMATION:

```

GENERAL INFORMATION:
APPLICANT:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

;	TITLE OF INVENTION:	CELLS
		10

; NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:

; CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Arnold Whit

ADDRESSEE: ARNOLD, WHITE
STREET: P.O. BOX 4433

STREET: P.O. BOX 4433
CITY: Houston

CITY: Houston
STATE: Texas

; COUNTRY: United States

; ZIP: 77210

; COMPUTER READABLE FORM:

```

;
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compat

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS

; OPERATING SYSTEM: FC D;
; SOFTWARE: PatentIn Release

```

;
; SOFTWARE: #1.30
;

```

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT

FILING DATE: CONCURRENT
CLASSIFICATION:

; CLASSIFICATION:

OM of: US-09-049-696-3 to: A_Geneseq_1101.* out_format : pfs

Date: Mar 30, 2002 2:17 AM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+ntp.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09049696/runat_28032002_145237_2001/app_query.fasta_1.12579
-DB=A_Geneseq_1101 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09049696 @CGNL_1_293 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-3

Query length: 240

Database: A_Geneseq_1101.*

Database sequences: 522463

Database length: 74073290

Search time (sec): 819.910000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AA675614				442.00	1055.12
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AA674824				442.00	1054.60
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AA674733				442.00	1054.60
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AA675474				442.00	1054.47
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AA624514				439.00	1047.20
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AA673716				439.00	1047.20
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AA674822				356.00	842.63
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AA673715				356.00	842.63
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA676749				306.00	719.32
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AA675660				306.00	719.32
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AA665272				306.00	719.32
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AA641039				239.50	559.98
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA611324				239.50	559.98
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AA670459				239.50	556.97
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA611325				239.50	556.97
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA611325				239.50	556.97
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AA674823				239.50	555.14
/SID22/gcgdata/geneseq/geneseq/AA1999.DAT:AA641036				239.50	555.14
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA611321				239.50	555.14
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA631800				73.00	152.29
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA631799				73.00	152.06
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA631798				73.00	151.97
/SID22/gcgdata/geneseq/geneseq/AA1989.DAT:AA691223				67.50	138.89
/SID22/gcgdata/geneseq/geneseq/AA1996.DAT:AA695268				66.00	126.96
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AA606666				66.00	126.96
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AA693284				64.00	126.51
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA694960				64.00	124.94
/SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AA694811				64.00	124.32
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA651208				63.50	131.65
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA651207				63.50	131.53
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA651206				63.50	130.31
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA658507				63.00	139.53
/SID22/gcgdata/geneseq/geneseq/AA1986.DAT:AA606665				63.00	133.96
/SID22/gcgdata/geneseq/geneseq/AA1986.DAT:AA606666				63.00	133.96
/SID22/gcgdata/geneseq/geneseq/AA1986.DAT:AA606441				63.00	133.96
/SID22/gcgdata/geneseq/geneseq/AA1987.DAT:AA606442				63.00	133.96
/SID22/gcgdata/geneseq/geneseq/AA1988.DAT:AA606443				63.00	133.96
/SID22/gcgdata/geneseq/geneseq/AA1990.DAT:AA604215				63.00	133.96
/SID22/gcgdata/geneseq/geneseq/AA1990.DAT:AA604216				63.00	133.96

/SID22/gcgdata/geneseq/geneseq/AA1990.DAT:AA606331 + 63.00 133.96 6.99 248
/SID22/gcgdata/geneseq/geneseq/AA1986.DAT:AA606661 + 63.00 133.04 7.19 271
/SID22/gcgdata/geneseq/geneseq/AA1990.DAT:AA604212 + 63.00 133.04 7.19 271
/SID22/gcgdata/geneseq/geneseq/AA1990.DAT:AA604217 + 63.00 133.04 7.19 271
/SID22/gcgdata/geneseq/geneseq/AA2000.DAT:AA649306 - 63.00 129.56 8.04 379

seq_name: /SID22/gcgdata/geneseq/geneseq/AA2001.DAT:AA675614

seq_documentation_block:

ID AAG75614 standard; Protein; 869 AA.

XX AAG75614;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:6378.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

XX colorectal carcinoma.

XX Homo sapiens;

XX WO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

XX 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI, 2001-235357/24.

XX N-PSDB; AAH35019.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers - Claim 11; Page 7851-7854; 9803pp; English.

XX AAH32943 to AAH37195 and AAG7788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patients own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAG7788 represent sequences used in the exemplification of the present invention.

XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 869 AA;

alignment_scores:

Quality: 442.00 Length: 79
Ratio: 5.595 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-3 x AAG75614

Align seg 1/1 to: AAG75614 from: 1 to: 869

3 AATGCTGATGTCGTTGCTGCTACTCTCCAGGTAATGATGAACC 52
|||||
56 AsnAlaaspValLeuValAlaGluSerThrProGlyAsnAspGluPr 72
|||||
53 CTACACTGACGAGATGGCAACTGTGGAGAGAGAGGTTGAAGGATCCACC 102
|||||
72 oTyThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgileHisL 89
|||||
103 TCATCTCTGATTTTCATTCGACGAGAAAAAGTTAGTGAATATGGACCAAA 152
|||||
89 euThrProAspPheileAlaGlyLysLysLeuAlaGluTyrglyProGln 105
|||||
153 GGTAGGGCATTTGTCATGACGAGGCTCATCTACGATGGGGAGTATTGA 202
|||||
106 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 122
|||||
203 CGAGTACAAATATGATGAGAAATTTCTACTTATCCAAT 239
|||||
122 pGluTyrrAsnAsnAspGluLysPheTyrrLeuSerAsn 134
|||||

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT.AAB74824

seq_documentation_block:

ID AAB74824 standard; Protein: 914 AA.

AC AAB74824;

XX 13-JUN-2001 (first entry)

XX Human ICACC-1 protein sequence.

XX ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW inflammatory bowel disease; autoimmune disease.

XX Homo sapiens.

XX WO9944620-A1.

XX 10-SEP-1999.

XX 03-MAR-1999; 99WO-US04703.

XX 03-MAR-1998; 98US-0076815.

XX (MAGA-) MAGAININ PHARM INC.

XX Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;

XX Nicolaides NC, Zhou Y, Dong Q;

XX WPT: 1999-550979/46.

XX N-PSDB; AAF81927.

XX New nucleic acid encoding calcium activated chloride channel, used to
PT identify, e.g. specific modulators for treating atopic allergy -
PS Claim 11; Fig 4B; 75pp; English.

XX The present sequence represents the human interleukin 9 (IL-9) induced
CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
CC have anti-allergic, anti-asthmatic, anti-inflammatory and
CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
CC used to alleviate asthma (or more generally atopic allergy), while those
CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
CC (increased levels) or IBD (reduced levels), also for monitoring
CC treatment of these conditions. The ICACC proteins can be used:

CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
CC reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
CC identify modulators and binding partners. ICACC polynucleotides can be
CC used to generate transgenic animals or recombinant cells, used to screen
CC for antagonists, also as a source of therapeutic antisense agents or
CC diagnostic probes (for quantifying mRNA expression, e.g. for
XX identification of modulators).

SQ Sequence 914 AA;

alignment_scores:

Quality: 442.00 Length: 79
Ratio: 5.595 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-3 x AAB74824

Align seg 1/1 to: AAB74824 from: 1 to: 914

3 AATGCTGATGTCGTTGCTGCTACTCTCCAGGTAATGATGAACC 52
|||||
101 AsnAlaaspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
|||||
53 CTACACTGACGAGATGGCAACTGTGGAGAGAGGTTGAAGGATCCACC 102
|||||
117 oTyThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgileHisL 134
|||||
103 TCATCTCTGATTTTCATTCGACGAGAAAAAGTTAGTGAATATGGACCAAA 152
|||||
134 euThrProAspPheileAlaGlyLysLysLeuAlaGluTyrglyProGln 150
|||||
153 GGTAGGGCATTTGTCATGACGAGGCTCATCTACGATGGGGAGTATTGA 202
|||||
151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
|||||
203 CGAGTACAAATATGATGAGAAATTTCTACTTATCCAAT 239
|||||
167 pGluTyrrAsnAsnAspGluLysPheTyrrLeuSerAsn 179
|||||

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2001.DAT.AAB74733

seq_documentation_block:

ID AAB74733 standard; Protein: 914 AA.

AC AAB74733;

XX 12-JUN-2001 (first entry)

XX Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.

XX Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
KW dermatological; immunosuppressive; anti-inflammatory; anti-HIV;
KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
KW antialzheimer; antiparkinsonian; antimicrobial; vulnery; gene therapy;
KW immune disorder; hyperproliferative disorder; cardiovascular disease;
KW cancer; angiogenic disorder; neurological disorder; infectious disease;
KW wound healing; regeneration; chemotaxis; chromosome 1.

XX Homo sapiens.

XX WO200112775-A2.

XX 22-FEB-2001.

XX 16-AUG-2000; 2000WO-US22325.

XX 17-AUG-1999; 99US-0149182.

XX (HUMA-) HUMAN GENOME SCI INC.

```
XX Rosen CA, Ni J, Florence KA, Fiscella M, Wei P, Baker KP;
PI Birse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;
XX
XX WPI; 2001-147550/15.
XX N-PSDB; AAF81787.
XX
XX Nucleic acids encoding 25 human secreted polypeptides, useful for
XX preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX disease and diabetic retinopathy -
XX
XX Claim 11; Page 459-460; 485pp; English.
XX
XX AAF81787 to AAF81817 encode the human secreted proteins given in AAF74733
XX to AAF74772. Human secreted proteins can have activities based on the
XX tissues and cells they are expressed in. Example of activities include:
XX immunomodulatory; antisclerotic; dermatologic; immunosuppressive;
XX antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiac;
XX vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic;
XX anticonvulsant; antialzheimers; antiparkinsonian; antimicrobial; and
XX vaccine. Human secreted proteins can be used in gene therapy and
XX (PEP1) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate polypeptide expression. For example, NAM1
XX and PEP1 may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patients genome
XX that affect the activity of proteins by expressing inactive proteins or
XX to supplement the patients own production of polypeptides. Disorders that
XX may be prevented, diagnosed and/or treated include immune disorders,
XX hyperproliferative disorders (e.g. cancers), cardiovascular diseases,
XX angiogenic disorders, neurological disorders, infectious diseases and/or
XX for promoting wound healing, regeneration and/or chemotaxis. AAF81778 to
XX AAF81786 and AAF74732 represent sequences used in the exemplification of
XX the present invention.
XX
XX Sequence 914 AA;
SQ
alignment_scores:
Quality: 442.00 Length: 79
Ratio: 5.595 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-049-696-3 x AAF74733
Align seq 1/1 to: AAF74733 from: 1 to: 914
3 AATGCTCATGTTCTGCTGCTGAGTCTACTCTCCAGGTAATGATGAACC 52
101 AsnAlaAspValLeuValAlaGluSerThrProProGlyAsnAspGluPr 117
53 CTACACTGAGCAGATGGCAACTGTGTGAGAGAAGGGTGAAGGATCCACC 102
117 OTYrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134
103 TCACCTCTGATTTTCATTGTCAGGAAAAAAGTTAGCTGAATGAGCACAA 152
134 eutrProAspPheIleAlaGlyLysLysLeuAlaGluTyGlyProGln 150
153 GGTAGGCGATTTCCTCATGATGGGCTCATCTACGATGGGAGCTATTGA 202
151 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
203 CGAGTACAATAATGATGAGAAATCTACTTATCCAAAT 239
167 pGluTyAsnAsnAspGluLysPheTyLeuSerAsn 179
seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAF75474
seq_documentation_block:
ID AAG75474 standard; Protein; 925 AA.
XX
```

```
AC AAG75474;
XX
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen protein SEQ ID NO:6238.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; chromosome 1.
XX
XX Homo sapiens;
XX
XX WO200122920-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
XX 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI; 2001-235357/24.
XX N-PSDB; AAF34879.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 11; Page 7686-7690; 9803pp; English.
XX
XX AAF32943 to AAF37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAF37196 to AAF37204
XX and AAF77789 represent sequences used in the exemplification of the
XX present invention.
XX
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 925 AA;
SQ
alignment_scores:
Quality: 442.00 Length: 79
Ratio: 5.595 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-049-696-3 x AAG75474
Align seq 1/1 to: AAG75474 from: 1 to: 925
3 AATGCTCATGTTCTGCTGCTGAGTCTACTCTCCAGGTAATGATGAACC 52
112 AsnAlaAspValLeuValAlaGluSerThrProProGlyAsnAspGluPr 128
53 CTACACTGAGCAGATGGCAACTGTGTGAGAGAAGGGTGAAGGATCCACC 102
123 OTYrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 145
```

103 TCACCTCCTGATTTCACGAGAAAAGTTAGCTGATATGGACCAAA 152
|||||
145 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 161
|||||
153 GGTAGGGGATTGTCATGAGTGGGCTCATCTACGATGGGGAGTATTGA 202
|||||
162 GlyArgAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 178

203 CGAGTACAATAATGATGAGAAATCTACTTATCCAAAT 239
|||||
178 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsn 190
|||||

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA24514

seq_documentation_block:
ID AAM24514 standard; Protein: 914 AA.

XX AC AAM24514;
XX 12-OCT-2001 (first entry)
XX C902P predicted amino acid sequence.

XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW immunogenic; gene therapy; vaccine; colonic cancer.
XX OS Homo sapiens.

XX WO200149716-A2.

XX 12-JUL-2001.

XX 29-DEC-2000; 2000WO-US35596.

XX 30-DEC-1999; 99US-0476296.

XX 10-JAN-2000; 2000US-0480321.

XX 15-FEB-2000; 2000US-0504629.

XX 06-MAR-2000; 2000US-0519444.

XX 19-MAY-2000; 2000US-0575251.

XX 29-JUN-2000; 2000US-0609448.

XX 28-AUG-2000; 2000US-0649811.

XX (CORI-) CORIXA CORP.

XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;

XX King GE, Wang T, Jiang Y;

XX WPI; 2001-441847/47.

XX Colon tumor associated proteins and nucleic acids useful for the

XX prevention, diagnosis and treatment of colonic cancer.

XX Claim 2; Page 440-443; 472pp; English.

XX The present invention describes colon tumour associated proteins (I) and
XX the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX (II) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate colon tumour associated protein (TCAP).
XX expression, such as colonic cancer. For example, (I) and (II) may be
XX used to treat disorders associated with decreased expression by
XX rectifying mutations or deletions in a patient's genome that affect the
XX activity of TCAPs by expressing inactive proteins or to supplement the
XX patients own production of them. Additionally, (II) may be used to
XX produce the TCAP proteins, by inserting the nucleic acids into a host
XX cell culturing the cell to express the protein. (II) and its
XX complementary sequences may also be used as DNA probes in diagnostic
XX polymerase chain reaction (PCR) and hybridisation assays to detect and
XX quantitate the presence of similar nucleic acids in samples, and
XX therefore which patients may be in need of restorative therapy. (I) may
XX also be used as antigens in the production of antibodies against TCAPs
XX and in assays to identify modulators of TCAP expression and activity.
XX Anti-(I) antibodies and antagonists may also be used to down regulate

CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.

XX Sequence 914 AA;

alignment_scores:
Quality: 439.00 Length: 79
Ratio: 5.557 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.734

alignment_block:

US-09-049-696-3 x AAM24514 ..

Align seg 1/1 to: AAM24514 from: 1 to: 914

3 AATGCTGATGTTCTGCTGCTGCTACTCTCCAGGTAATGATGAACC 52
|||||
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
|||||
53 CTACACTGACGATGGCGAACTGTGTGAGAGAAAGGTGAAAGGATCCACC 102
|||||
117 cTyrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134
|||||
103 TCACCTCCTGATTTCATGTCAGGAGAAAGTTAGCTGAATATGGACCAAA 152
|||||
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyrGlyProGln 150
|||||
153 GGTAGGGCATTGTCCATGATGGGCTCATCTACGATGGGGAGTATTGA 202
|||||
151 GlyLysAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
|||||
203 CGAGTACAATAATGATGAGAAATCTACTTATCCAAAT 239
|||||
167 pGluTyrAsnAsnAspGluLysPheTyrLeuSerAsn 179
|||||

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA273716

seq_documentation_block:

ID AAB73716 standard; Protein: 914 AA.

XX AAB73716;

XX 11-SEP-2001 (first entry)

XX Human CLCA1 protein, SEQ ID NO:2.

XX Human CLCA1; goblet cell; mouse Gob-5 orthologue; drug screening;
XX expression inhibition; antisense therapy; gene therapy;
XX chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Misc-difference 152 /note= "Encoded by AGG in AAB46124"

XX WO200138530-A1.

XX 31-MAY-2001.

XX 22-NOV-2000; 2000WO-JP08232.

XX 24-NOV-1999; 99JP-0333479.

XX 27-APR-2000; 2000JP-0127589.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Nakanishi A, Morita S;

DR WPI: 2001-355935/37.
DR N-PSDB; AAH46102, AAH46124.
XX
PT New antisense nucleotide, useful for treatment and prevention of
PT bronchial asthma and chronic obstructive pulmonary disease
XX
PS
PS Claim 2; Page 76-80; 10app; Japanese.
XX
CC The invention relates to an antisense nucleotide targetted to the mouse
CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
CC the CLCA1 gene (coding sequence shown in AAH46102). The invention also
CC relates to an antibody specific for the Gob-5 protein, medical and
CC diagnostic compositions containing the antisense nucleotide or the
CC antibody, and methods and kits for screening for compounds which inhibit
CC the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.
CC The antisense oligonucleotides and antibody are therefore useful for the
CC treatment and prevention of bronchial asthma and chronic obstructive
CC pulmonary disease. The present sequence represents human CLCA1 protein.
XX
SQ Sequence 914 AA;

alignment_scores:
Quality: 439.00 Length: 79
Ratio: 5.557 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.734

alignment_block:
US-09-049-696-3 x AAB73716 ..

Align seg 1/1 to: AAB73716 from: 1 to: 914

3 AATGCTGATGTTCTGTTGCTGAGTCTACTCTCCAGGTAATGATGAACC 52
|||||
101 AsnAlaAspValLeuValAlaGluSerThrProGlyAsnAspGluPr 117
|||||
53 CTACACTGACGATGGCACTGTGGAGAGAAGGGTGAAGGATCCACC 102
|||||
117 OTYrThrGluGlnMetGlyAsnCysGlyGluLysGlyGluArgIleHisL 134
|||||
103 TCACCTCTGATTTTCATTCGAGGAAAAGTTAGCTGAATATGACCAAA 152
|||||
134 euThrProAspPheIleAlaGlyLysLysLeuAlaGluTyGlyProGln 150
|||||
153 GGTAGGCGATTTGTCATGAGTGGGCTCATCTACGATGGGCGAGTATTGA 202
|||||
151 GlyLysAlaPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 167
|||||
203 CGAGTACAATAATGATGAGAAATTCCTACTTATCCAAT 239
|||||
167 pGluTyAsnAsnAspGluLysPheTyLeuSerAsn 179
|||||

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74822

seq_documentation_block:
ID AAB74822 standard; Protein; 913 AA.
XX
AC AAB74822;
XX
DT 13-JUN-2001 (first entry)
XX
DE Murine ICACC-1 protein sequence.
XX
KW ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW inflammatory bowel disease; autoimmune disease.
XX
OS Mus sp.
XX
PN W09944620-A1.
XX

PD 10-SEP-1999.
XX
PF 03-MAR-1999; 99WO-US04703.
XX
PR 03-MAR-1998; 98US-0076815.
XX
RA (MAGA-) MAGA IN PHARM INC.
XX
PI Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;
PI Nicolaides NC, Zhou Y, Dong Q;
XX
DR WPI: 1999-550979/46.
DR N-PSDB; AAF81925.
XX
PT New nucleic acid encoding calcium activated chloride channel, used to
PT identify, e.g. specific modulators for treating atopic allergy
XX
PS Claim 12; Fig 2; 75pp; English.
XX
CC The present sequence represents the murine interleukin 9 (IL-9) induced
CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
CC have anti-allergic, anti-asthmatic, anti-inflammatory and
CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
CC used to alleviate asthma (or more generally atopic allergy), while those
CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
CC (increased levels) or IBD (reduced levels), also for monitoring
CC treatment of these conditions. The ICACC proteins can be used:
CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
CC reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
CC identify modulators and binding partners. ICACC polynucleotides can be
CC used to generate transgenic animals or recombinant cells, used to screen
CC for antagonists, also as a source of therapeutic antisense agents or
CC diagnostic probes (for quantifying mRNA expression, e.g. for
CC identification of modulators).
XX
SQ Sequence 913 AA;

alignment_scores:
Quality: 356.00 Length: 78
Ratio: 4.877 Gaps: 0
Percent Similarity: 93.590 Percent Identity: 79.487

alignment_block:
US-09-049-696-3 x AAB74822 ..

Align seg 1/1 to: AAB74822 from: 1 to: 913

3 AATGCTGATGTTCTGTTGCTGAGTCTACTCTCCAGGTAATGATGAACC 52
|||||
101 AsnAlaAspValLeuValSerThrThrSerProLeuGlyAsnAspGluPr 117
|||||
53 CTACACTGACGATGGCACTGTGGAGAGAAGGGTGAAGGATCCACC 102
|||||
117 OTYrThrGluHisIleGlyAlaCysGlyGluLysGlyIleArgIleHisL 134
|||||
103 TCACCTCTGATTTTCATTCGAGGAAAAGTTAGCTGAATATGACCAAA 152
|||||
134 euThrProAspPheLeuAlaGlyLysLysLeuThrGlnTyGlyProGln 150
|||||
153 GGTAGGCGATTTGTCATGAGTGGGCTCATCTACGATGGGCGAGTATTGA 202
|||||
151 AspArgThrPheValHisGluTrpAlaHisPheArgTrpGlyValPheAs 167
|||||
203 CGAGTACAATAATGATGAGAAATTCCTACTTATCCAAT 236
|||||
167 nGluTyAsnAsnAspGluLysPheTyLeuSer 178
|||||

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB73715

150 GlyLysLeuPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 166
203 CGAGTACAAATATGATGAGAAATCTTAC 230
166 pGluTrpAsnGluAspGlnProPheTyr 175

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB87560

seq_documentation_block:
ID AAB87560 standard; Protein; 919 AA.

AC AAB87560;
DT 15-MAY-2001 (first entry)
DE Human PRO1124.
KW Human; PRO protein; mapping.
XX
OS Homo sapiens.
PN WO200116318-A2.
XX
PD 08-MAR-2001.
XX
PF 24-AUG-2000; 2000WO-US233328.
XX
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21030.
PR 07-DEC-1999; 99US-0169495.
PR 09-DEC-1999; 99US-0170262.
PR 11-JAN-2000; 2000US-0175481.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 01-MAR-2000; 2000WO-US05601.
PR 03-MAR-2000; 2000US-0187202.
PR 25-APR-2000; 2000US-0199397.
PR 22-MAY-2000; 2000WO-US14042.
PR 05-JUN-2000; 2000US-0209832.
XX
PA (GETH) GENENTECH INC.
XX
PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
XX
DR WPI; 2001-183260/18.
DR N-PSDB; AAF92092.
XX
PT Eighty four nucleic acids encoding PRO polypeptides, useful in
PT molecular biology, including use as hybridization probes, and in
PT chromosome and gene mapping.
XX
PS Claim 12; Fig 70; 278pp; English.
XX
SS The present sequence is a human PRO polypeptide (secreted and
CC transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
CC anti-PRO antibodies are useful for preparation of a medicament useful in
CC the treatment of a condition which is responsive to the PRO protein,
CC agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
CC employed as molecular weight markers for protein electrophoresis. The PRO
CC coding sequence has applications in molecular biology, including use as
CC hybridisation probes, and in chromosome and gene mapping.
XX
SQ Sequence 919 AA;

alignment_scores:
Quality: 306.00 Length: 76
Ratio: 4.708 Gaps: 0
Percent Similarity: 85.526 Percent Identity: 69.737

alignment_block:
US-09-049-696-3 x AAB87560
Align_seg 1/1 to: AAB87560 from: 1 to: 919
3 AATGCTGATGTTCTGTTGCTGAGTCTCTCTCCAGGTAATGATGAACC 52
100 HisAlaAspValIleValAlaProThrLeuProGlyArgAspGluPr 116
53 CTACACTGAGCAGATGGCACTGTGGAGAGAGGTGAAGAGTCCACC 102
116 oTyThrLysGlnPheThrGluCysGlyGluLysGlyGluTyrIleHisP 133
103 TCACTCTGATTTTCAATTGTCAGGAAAAAGTTAGCTGAATATGACCAAA 152
133 heThrProAspLeuLeuGlyLysLysGlnAsnGluTyrGlyProPro 149
153 GGTAGGCAATTTGTCATGAGTGGGCTCATCTACGATGGGAGTATTGA 202
150 GlyLysLeuPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 166
203 CGAGTACAAATATGATGAGAAATCTTAC 230
166 pGluTrpAsnGluAspGlnProPheTyr 175

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB65272

seq_documentation_block:
ID AAB65272 standard; Protein; 919 AA.

AC AAB65272;
DT 02-APR-2001 (first entry)
DE Human PRO1124 (UNQ562) protein sequence SEQ ID NO:379.
XX
KW Human; secreted and transmembrane protein; PRO; cytostatic;
KW cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
XX diagnostic assay.
XX
OS Homo sapiens.
XX
PN WO2000073454-A1.
XX
PD 07-DEC-2000.
XX
PF 30-MAR-2000; 2000WO-US08439.
XX
PR 02-JUN-1999; 99WO-US12252.
PR 23-JUN-1999; 99US-0141037.
PR 07-JUL-1999; 99US-0143048.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 28-JUL-1999; 99US-0146222.
PR 17-AUG-1999; 99US-0149396.
PR 15-SEP-1999; 99WO-US21090.
PR 08-OCT-1999; 99US-0158663.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28301.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 05-JAN-2000; 2000WO-US00219.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 22-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 02-MAR-2000; 2000WO-US05841.
PR 15-MAR-2000; 2000WO-US06884.
PR 20-MAR-2000; 2000WO-US07377.
XX

```

PA (GETH ) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Kljavin IJ, Napier MA, Pan J, Paoni NF;
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WT;
PI Zhang Z;
XX
DR WPI: 2001-032160/04.
DR N-PSDB: AAF44241.
XX
PT PRO polynucleotides used to produce polypeptides used to target
PT bioactive molecules such as toxins, radiolabels or antibodies, to
PT specific cells, to cause targeted cell death -
XX
PS Claim 12; Fig 274; 935pp; English.
XX
CC The present invention describes human secreted and transmembrane PRO
CC proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC can be used for targeted delivery of bioactive molecules, such as
CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC sequences, and their fragments, can be used as hybridisation probes, in
CC chromosomal and gene mapping, and in the generation of anti-sense RNA
CC and DNA. They may also be used to produce transgenic animals which are
CC used to develop and screen therapeutically useful reagents. The PRO
CC nucleotide and protein sequence can be used for tissue typing and in
CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
CC in the isolation of human PRO sequences. AAF44087 to AAF44269 and
CC AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC sequences given in the exemplification of the present invention.
XX
SQ Sequence 919 AA:

alignment_scores:
  Quality: 306.00 Length: 76
  Ratio: 4.708 Gaps: 0
Percent Similarity: 85.526 Percent Identity: 69.737

alignment_block:
US-09-049-696-3 x AAB65272 ..

Align seg 1/1 to: AAB65272 from: 1 to: 919

3 AATGCTGATGTTCTGCTGCTAGTCTACTCTCCAGGTAATGATGAACC 52
  :::::::::::::::::::: ||| ::::::::::::::::::::
100 HisAlaAspValIleValAlaProProThrLeuProGlyArgAspGluPr 116

53 CTACACTGACGACATGGCAACTGTGGAGAGAGGCTGAAGGATCCACC 102
  :::::::::::::::::::: ::::::::::::::::::::
116 ofYrThrLysGlnPheThrGluCysGlyGluLysGlyGluTyrlleHisP 133

103 TCACTCCTCATTCATTCAGTGGAGAGAGGCTGAAGGATCCACC 152
  :::::::::::::::::::: ::::::::::::::::::::
133 heThrProAspLeuLeuLeuGlyLysGlnAsnGluTyrlleHisP 149

153 GGTAGGCGCATTTGTCATGAGTGGGCTCATCTACGATGGGAGTATTGA 202
  :::::::::::::::::::: ::::::::::::::::::::
150 GlyLysLeuPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 166

203 CGAGTACAATAATGATGACAAATTCATC 230
  :::::::::::::::::::: ::::::::::::::::::::
166 pGluTyrlleAsnGluAspGlnProPheTyrl 175

seq_name: /SID52/seqdata/geneseq/geneseq/AA1999.DAT:AAV41039
seq_documentation_block:
ID AAV41039 standard; protein; 592 AA.
XX
AC AAV41039;
XX
DT 07-DEC-1999 (first entry)

```

```

XX Human lung tumor antigen L762P variant 1.
DE
XX Human; lung tumor; lung cancer; T cell stimulation.
KW
XX Homo sapiens.
OS
XX MO9947674-A2;
XX 23-SEP-1999.
PD
XX 17-MAR-1999; 99WO-US05798.
XX 18-MAR-1998; 98US-0040802.
XX 18-MAR-1998; 98US-0040984.
XX 27-JUL-1998; 98US-0123912.
XX 27-JUL-1998; 98US-0123933.
XX (CORI-) CORIXA CORP.
XX Read SG, Wang T;
XX WPI: 1999-571839/48.
XX N-PSDB: AAZ24657.
XX New isolated lung tumor polynucleotides, used to develop products for
XX the treatment, prevention and monitoring the progression of lung cancer
XX Example 3; Page 144-145; 148pp; English.
XX
CC The invention provides isolated human lung tumor nucleic acids and
CC polypeptides. The polypeptides can be used for the treatment of lung
CC cancer. The polypeptides and polynucleotides can be used to stimulate T
CC cells or antigen presenting cells for use in the treatment of lung
CC cancer. The polypeptides and monoclonal antibodies specific for the
CC polypeptides can also be used to inhibit the development of lung cancer.
CC Agents which bind the polypeptides can be used for detecting lung cancer
CC and for monitoring the progression of lung cancer.
XX
SQ Sequence 592 AA:

alignment_scores:
  Quality: 239.50 Length: 78
  Ratio: 3.802 Gaps: 1
Percent Similarity: 80.769 Percent Identity: 55.128

alignment_block:
US-09-049-696-3 x AAY41039 ..

Align seg 1/1 to: AAY41039 from: 1 to: 592

5 GCTGATGTTCTGTTGCTGAGTCTACTCTCCAGGTAATGATGAACCCCTA 55
  :::::::::::::::::::: ::::::::::::::::::::
109 AlaAsnValIleValThrAspTrpTyrlleHisGlyAspPheArg 125

56 CACTGAGCAGATGGCAACTGTGGAGAGAGGCTGAAGGATCCACCTCA 105
  :::::::::::::::::::: ::::::::::::::::::::
125 rThrLeuGlnTyrlleArgGlyCysGlyLysGluGlyLysTyrlleHisPhe 142

106 CTCCTCATTTTCATTCGAGGAGAGAGAGTGA...CTCAATATGGACACAA 152
  :::::::::::::::::::: ::::::::::::::::::::
142 hrProAsnPheLeuLeuAsnAspAsnLeuThrAlaGlyTyrlleArg 158

153 GGTAGGCGCATTTGTCATGAGTGGGCTCATCTACGATGGGAGTATTGA 202
  :::::::::::::::::::: ::::::::::::::::::::
159 GlyArgValPheValHisGluTrpAlaHisLeuArgTrpGlyValPheAs 175

203 CGAGTACAATAATGATGACAAATTCATCTATCC 236
  :::::::::::::::::::: ::::::::::::::::::::
175 pGluTyrlleAsnAspAspPheProPheTyrlleAsn 186

```

I25 rInrLeugmIYrAlgerYCyscrfJyscrac-fJsf-J frrrem---

align "seq 1/1" to: AAY41040 from: 1 to: 791

6 GCTGATGTTCTCGTTCGTGACTACTCTCCAGGTAATGATGAACCTTA 55
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 109 AlaAsnValIleValThrAspTrpTyrGlyAlaHisHisAspAspProTy 125
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 56 CACTGACGAGATGGCCAACTCTGGAGAGAAGGTGAAGGATCCACCTCA 105
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 125 rThrLeuGlnTyrArgGlyCysGlyLysGluGlyLysTyrIleHisPhe⁺ 142
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 106 CTCCTGATTTTCATTCGACGAAAAAAGTTA...GCTGAATATGGACACAA 152
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 142 hrProAsnPheLeuLeuAsnAspAsnLeuThrAlaGlyTyrGlySerArg 158
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 153 GTTAGGGCATTTTGTCATGACTGGGCTCATCTACGATGGGAGCATTTTCA 202
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 159 GlyArgValPheValHisLeuTrpAlaHisLeuArgTrpGlyValPheAs 175
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 203 CGAGTACAATAATGATGAGAAATTTCTACTTATCC 236
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 175 pGluTyrAsnAsnAspLysProPheTyrIleAsn 186
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

seq_name: /SIDS2/gcgcdata/geneseq/geneseq/AA2000.DAT: AAB11325

seq_documentation_block:

Doc-Documentation_Block.
ID AAB11325 standard; Protein; 791 AA.

AA AAB11325:
AC

DT 21-FEB-2001 (first entry)

DE Human lung cancer-associated protein L762P variant 2.

Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic; vaccine; detection.

OS Homo sapiens.

XX PN WO200061612-A2.

PD 19-OCT-2000.

03-APR-2000: 2000WO-US08896.

XX
PR
02-APR-1999: 99US-028547

PR	1/-DEC-1999;	99US-046635
PR	30-DEC-1999:	99US-047649

PR 10-JAN-2000; 2000US-0480884.
PR 22-FEB-2000; 2000US-0510376.

XX
PA (CORI-) CORIXA CORP.

XX PI Wang T. Fan L:

XX
DR WPT: 2000-628399/60.

DR N-PSDB; AAC65897.
XX

PT isolated polypeptide comprising an immunogenic portion of a lung tumor protein is used for detecting and monitoring progression of lung cancer in a patient -

XX
PS
Claim 3: Page 178-180: 261pp: English-

This invention describes a novel isolated polypeptide (I) which comprising an immunogenic portion of a lung tumor protein or variant (P2) which have cytostatic activity. The polypeptides and polynucleotides are used in compositions and vaccines to inhibit the development of cancer, especially lung cancer, in a patient. Methods described in the invention can be used to monitor the progression of a cancer by carrying out the detection at subsequent time points and comparing the results from the different time points. CD4+ and/or CD8+ T-Cells isolated from a patient are treated with P2, polynucleotides encoding P2 or antigen presenting cells expressing P2 and then administered to the patient to inhibit development of cancer.

us-09-049-696-3.rag

This Page Blank (uspto)

OM of: US-09-049-696-2 to: SPTRMBL_17.* out_format : pfs
Date: Mar 30, 2002 2:46 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q/cgn2_1/USPRO.spool/US09049696/runat_28032002_145238_2053/app_query.fasta_1.12579
-DB=SPTRMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62
-TRANS=human40.cdd -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 @CGN1_1_705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-049-696-2
Query length: 273
Database: SPTRMBL_17.*
Database sequences: 473505
Database length: 146272329
Search time (sec): 805.760000

score_list:	Strd	Orig	ZScore	EScore	Len	Documentation
sp_mammal:Q9TUB5	+	382.00	803.49	1.9e-37	917	! Q9TUB5 sus scrofa (pig). epithel
sp_rodent:O88826	+	371.00	779.77	4.0e-36	913	! O88826 mus musculus (mouse). g
sp_rodent:Q9D726	+	371.00	779.77	4.0e-36	913	! Q9D726 mus musculus (mouse). ch
sp_human:Q9UNF7	+	314.00	656.68	2.9e-29	917	! Q9UNF7 homo sapiens (human). ca
sp_mammal:O18744	+	288.00	608.44	3.8e-26	942	! O18744 bos taurus (bovine). lu
sp_mammal:O18742	+	288.00	601.70	3.9e-26	794	! O18742 bos taurus (bovine). lu
sp_mammal:O18743	+	288.00	601.44	3.9e-26	820	! O18743 bos taurus (bovine). lu
sp_mammal:O18741	+	288.00	600.65	3.9e-26	905	! O18741 bos taurus (bovine). lu
sp_rodent:Q9R070	+	287.00	598.52	5.1e-26	902	! Q9R070 mus musculus (mouse). ca
sp_rodent:Q9E0R4	+	287.00	598.52	5.1e-26	902	! Q9E0R4 mus musculus (mouse). en
sp_rodent:O88860	+	285.00	594.21	8.9e-26	901	! O88860 mus musculus (mouse). ch
sp_rodent:Q9QX15	+	285.00	594.20	8.9e-26	902	! Q9QX15 mus musculus (mouse). ca
sp_human:Q9Y6N3	+	279.00	591.15	4.5e-25	262	! Q9Y6N3 homo sapiens (human). cl
sp_human:Q9UQC9	+	223.50	461.07	2.2e-18	943	! Q9UQC9 homo sapiens (human). ch
sp_human:Q9Y6N2	+	223.50	461.07	2.2e-18	943	! Q9Y6N2 homo sapiens (human). ca
sp_bacteria:Q9WXT8	+	81.00	154.17	0.3017	858	! Q9WXT8 thermotoga maritima. fla
sp_virus:Q71051	+	80.00	159.26	0.3883	347	! Q71051 human immunodeficiency v
sp_virus:Q9DV87	+	77.00	157.78	0.8762	186	! Q9DV87 human immunodeficiency v
sp_archaea:Q58006	+	75.50	144.77	1.37	630	! O58006 pyrococcus horikoshii. h
sp_virus:Q71097	+	71.00	139.58	4.70	358	! Q71097 human immunodeficiency v
sp_virus:Q90DF3	+	70.50	145.74	5.26	145	! Q9QDF3 human immunodeficiency v
sp_virus:Q71103	+	70.50	138.11	5.40	376	! Q71103 human immunodeficiency v
sp_virus:Q9QRM8	+	69.50	142.45	6.97	167	! Q9QRM8 human immunodeficiency v
sp_virus:Q75340	+	69.00	143.01	7.96	136	! Q75340 human immunodeficiency v
sp_archaea:Q54324	+	68.50	126.77	9.62	904	! Q54324 sulfolobus islandicus. c
sp_organella:Q9XKX6	-	68.00	138.88	10.57	174	! Q9XKX6 chrysemya picta. nadh de
sp_organella:P87935	+	68.00	132.86	10.79	369	! P87935 human immunodeficiency v
sp_virus:Q99C04	+	68.00	125.98	11.04	872	! Q99C04 human immunodeficiency v
sp_invertebrate:Q9UAY2	-	67.50	125.54	12.65	805	! Q9UAY2 caenorhabditis elegans
sp_plant:Q9XIF0	+	67.50	124.59	12.69	906	! Q9XIF0 arabidopsis thaliana (md
sp_virus:Q73161	+	67.00	134.77	14.04	222	! Q73161 human immunodeficiency v
sp_fungi:O13632	+	66.50	130.52	16.29	330	! O13632 schizosaccharomyces pom
sp_invertebrate:O17258	-	66.50	123.38	16.69	805	! O17258 caenorhabditis elegans
sp_invertebrate:O45152	-	66.50	123.38	16.69	805	! O45152 caenorhabditis elegans
sp_virus:Q70538	+	66.00	133.33	18.47	203	! Q70538 human immunodeficiency v
sp_archaea:Q9V171	+	66.00	131.69	18.57	249	! Q9V171 pyrococcus abyssi. atp-b
sp_organella:Q33549	+	66.00	130.33	18.64	288	! Q33549 blastocystidia culicis
sp_virus:Q71634	+	66.00	126.32	18.91	487	! Q71634 human immunodeficiency v
sp_virus:Q92247	+	65.50	132.57	21.19	195	! Q92247 human immunodeficiency v

sp_virus:Q9YTY0 + 65.50 132.17 21.22 205 ! Q9YTY0 human immunodeficiency
sp_plant:O81295 - 65.50 132.13 21.22 206 ! O81295 arabidopsis thaliana
seq_name: sp_mammal:Q9TUB5
seq_documentation_block:
ID Q9TUB5 PRELIMINARY; PRT; 917 AA.
AC Q9TUB5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE EPIHELIAL CHLORIDE CHANNEL PROTEIN.
GN AEC.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID:9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ILEAL MUCOSA;
RA Gaspar K.J., Gabriel S.E., Racette K.J., Forsyth G.W.;
RT "A CDNA involved in porcine exocrine chloride conductance."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF095584; AAF00077.1;
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWFA; 1.
SQ SEQUENCE 917 AA; 100735 MW; 649F41944F7EAD19 CRC64;

alignment_scores:
Quality: 382.00 Length: 91
Ratio: 4.716 Gaps: 0
Percent Similarity: 89.011 Percent Identity: 79.121
alignment_block:
US-09-049-696-2 x Q9TUB5 ..
Align seg 1/1 to: Q9TUB5 from: 1 to: 917
1 GTTGCATCGACCCCAATGTCAGAGATGAACACTCATTCAACAAT 50
37 ILEAIIAASPProasnValProGluAspGluArgLeuIleGlnAsnII 53
51 AAAGGACATGGTCAACAGCATCTCTGTATCTGTGTGAAGCTACAGAA 100
53 eLysAspMetValThrLysAlaSerProTyrLeuPheGluAlaThrGluL 70
101 AGCGATTTTATTTCAAAAATGTTGCCATTTTGTATTCCTGAACATGGAAG 150
70 ysArgPheTyrPheLysAsnValAlaIleLeuIleProAlaSerTrpLys 86
151 ACAAGGNTGACTATGTGACACCAAAACITGACACCTACAAAATGCTGA 200
87 AlaLysProGluTyrValLysProLysLeuGluThrTyrLysAsnAlaAs 103
201 TGTTCGTGTTGCTGAGTCTATCTCCAGGNAATGATGAACCTACACTG 250
103 pValValThrGluProAsnProGluAsnAspGlyProTyrThrG 120
251 NGCAGATGGCAACTGTGGCGAG 273
120 LucInMetGlyAsnCysGlyGlu 127
seq_name: sp_rodent:O88826
seq_documentation_block:
ID O88826 PRELIMINARY; PRT; 913 AA.
AC O88826;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

NCBI_taxID=10090;
RN [1]
OX
RN
NCBI_taxID=10090;
OX
RN [1]
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mus musculus (mouse).
OS


```

Ratio: 3.840      Gaps: 0
Percent Similarity: 82.418      Percent Identity: 59.341

alignment_block:
US-09-049-696-2 x Ol8742      ..

Align seg 1/1 to: Ol8742 from: 1 to: 794

```

1 GTTCCAAATCATCCCAATGTGCCAGAAAGATGAACAACTCATTTCAACAAT 50
36 ILeaIIeaLnProSerValProGLuAspGLuLeuIIeGLuAsnII 52
51 AAAGGACATGGTCACCCAGGCATCTCTGTATCTGTGTGAAGCTACAGGAA 100
52 eLYsGLuMeCValThrGLuAlaSerThrTyrLeuPheHisAlaThrLYsA 69

[illegible]

201 TGTTCTGGTTCGATCCTANTCCTCCAGNAATGATGAACCOCTACACGT 250
||| |||::: |||::: |||::: |||:::
102 pValIIEvalAAsnProTy rLeuLysTy rGlyAspAspProTy rThrL 119
||| |||::: |||::: |||::: |||:::

251 NGCAGATGGGCCAACTGTGCAG 273
||| |||::: |||::: |||::: |||:::
119 euGlnTy rGlyArgCysGlu 126

```
seq_name: sp_mammal:O18743
seq_documentation_block:
ID O18743
AC O18743; PRELIMINARY; PRT; 820 AA.
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DD 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
```

01-JUN-2001 (TREMBLRel. 17, Last annotation update)
LU-ECAM-1.
Bos taurus (Bovine).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Bos.
Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;
[1]
RN

RC
PC
DRP
SEQUENCE FROM N.A.
TISSUE=LUNG;
Elbie R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine F.
Goodwin A., Pauli B.U.;
RA Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
ERA Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001263; AA886531.1; -;
DR InterPro: IPR002035; vWFA;
DR PROSITE: PS50234; vWFA.1

```

DR      SMART; SM00327; VWA; 1.
SQ      SEQUENCE 820 AA; 91464 MW; 4E01330AEF5A1494 CRC64;

alignment_scores:
    Quality: 288.00
    Ratio: 3.840
    Percent Similarity: 82.418
    Gaps: 0
    Percent Identity: 59.341

```

```

alignment_block:
  US-09-049-696-2 x O18743      ..
  Align seg 1/1 to: O18743 from: 1 to: 820
  1 GTTGAATCGACCCCAATGTGCCAAGATGAAACACTCATTCACAAT 50
    .....

```

36 IleAlaIleAsnProSerValProGluAspGluLysLeuIleGluAsnIle 52

51 AAAGGACATGGTGACCCAGGCATCTCTGTATCTGTTTGAAGCTACAGGAA 100
52 eLysGluMetValThrGluAlaSerThrThrLeuPheHisAlaThrLysA 69
101 AGCGATTTTATTTCAAAATGTGGCATTTGTGATTCCTGAACATCGGAG 150
69 xGluValThrPheMetValThrGluAlaSerThrThrLeuPheHisAlaThrLysA 69

151 ACAAGGNTGACTATTGTGAGACCAAAACCTTGAGACCCTACAAAATGCCTGA 200
 :::~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
 85 SerLysSerGluTyrPheIleProLysGlnGluSerTyrAspClnAlaAs 102
 :::~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
 -201 TGTTCCTGGTTGCTGACTCTANTCCCTCCAGGNAAATGATGAACCTCACATCG 250
 :::~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
 162 GValIleValAlaIleArgMetCysLeuThrGlyGluGluGluGluGluGlu 200

```

251  NGCAGATGGGCAACTGTGGCGAG 273
    ||| |||:::|||||||
119  euGlnTyrGlyArgCysGlyGlu 126
seq.name: sp_mammal:O18741
seq_documentation_block:

```

	PRELIMINARY;	PRT; 905 AA.
018741	(TREMblrel. 05, Created)	
AC 018741;	(TREMblrel. 05, Last sequence update)	
AD 01-JAN-1998	(TREMblrel. 17, Last annotation update)	
DT 01-JAN-1998		
ET 01-JUN-2001		
LU-ECAM-1.		
DE Bos taurus (Bovine).		
OS Bos taurus (Bovine).		

CCOC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
NCBI_taxID=9913;
[1]
SEQUENCE FROM N.A.
TISSUE=LUNG;
Elbie R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine P.

Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.

EMBL:	AF001261; AA886529.1; -
InterPro:	IPRO02035; VWFA
InterPro:	IPRO03961; FN.III.
Pfam:	PF00041; fn3.1.
PROSITE:	PS50234; VWFA; 1.
SMART:	SM00327; VWFA; 1.

alignment_scores:	
Quality:	288.00
Ratio:	3.840
Percent Similarity:	82.418
Length:	91
Gaps:	0
Percent Identity:	59.341

```

US-09-049-696-2 x 018741      ..
Align seg 1/1 to: 018741 from: 1 to: 905

      1 GTTGCAGATCGACCCCAATGTGCCAGACATGAACACATCATTCACAACAAAT 50
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    36 IIEAlAlAlAeSnProSeVaIrProGluAspGluLysLeuIleGluAsnII 52

```

151 ACAAAAGGNTGACTATGTCGAGACCACAAACTTCAGACCTACAAAAATGCTGA 200

Ratio: 3.671 Gaps: 0
Percent Similarity: 83.516 Percent Identity: 58.242

alignment_block:

US-09-049-696-2 x Q9Y6N3

Align seg 1/1 to: Q9Y6N3 from: 1 to: 262

```

1 GTTGCATCGACCCCAATGTGCCAGAGATGAACACTATTCAACAAT 50
   ::::::::::::::::::::::::::::::::::::::::::::::
36 IleAlaIleAsnProSerValProGluAspGluLeuIleGlnAsn11 52
   ::::::::::::::::::::::::::::::::::::::::::::::
51 AAAGGACATGGTGACCCAGCATCTCTGTATCTGTGTTGAAGCTACAGGAA 100
   ::::::::::::::::::::::::::::::::::::::::::::::
52 eLysGluMetValThrGluAlaSerThrHisLeuPheHisAlaThrLysG 69
   ::::::::::::::::::::::::::::::::::::::::::::::
101 AGCGATTTTATTTCAAAAATGTTGCCATTTTGTATCTCTGAAACATGGAAG 150
   ::::::::::::::::::::::::::::::::::::::::::::::
69 InArgAlaTyPheArgAsnValSerIleLeuIleProMetThrTyLys 85
   ::::::::::::::::::::::::::::::::::::::::::::::
151 ACAAGAGTGCATGTATGTGACCAACAACTTGACACCTACAAAATGCTGA 200
   ::::::::::::::::::::::::::::::::::::::::::::::
86 SerLysSerGlnTyLeuIleProLysGlnGluThrTyAspGlnAlaAs 102
   ::::::::::::::::::::::::::::::::::::::::::::::
201 TGTCTGTTGTTGAGTCTANTCTCCAGGNAATGATGAACCTACACTG 250
   ::::::::::::::::::::::::::::::::::::::::::::::
102 pValIleValAlaAspLeuTyLeuLysTyGlyAspProTyThrL 119
   ::::::::::::::::::::::::::::::::::::::::::::::
251 NGCAGATGGGCAACTGTGGCGAG 273
   ::::::::::::::::::::::::::::::::::::::::::::::
119 euGlnTyGlyGlnCysGlyAsp 126

```

seq_name: sp_human:Q9UQC9

seq_documentation_block:

```

ID Q9UQC9 PRELIMINARY; PRT; 943 AA.
AC Q9UQC9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CHLORIDE CHANNEL PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-CORNEAL EPITHELIUM;
RA Itoh R., Kawamoto S., Kinoshita S., Kawasaki S., Okubo K.;
RT "Isolation and characterization of chloride channel in human corneal
epithelium.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026833; BAA77810.1; -.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWFA; 1.
SQ SEQUENCE 943 AA; 103939 MW; 4059766CFBD20E72 CRC64;

```

alignment_scores:

Quality: 223.50 Length: 91
Ratio: 3.020 Gaps: 1
Percent Similarity: 81.319 Percent Identity: 46.154

alignment_block:

US-09-049-696-2 x Q9UQC9

Align seg 1/1 to: Q9UQC9 from: 1 to: 943

```

1 GTTGCATCGACCCCAATGTGCCAGAGATGAACACTATTCAACAAT 50
   ::::::::::::::::::::::::::::::::::::::::::::::
45 IleAlaIleAsnProGlnValProGluAsnGlnAsnLeuIleSerAsn11 61

```

```

51 AAAGGACATGGTGACCCAGCATCTCTGTATCTGTGTTGAAGCTACAGGAA 100
   ::::::::::::::::::::::::::::::::::::::::::::::
61 eLysGluMetIleThrGluAlaSerPheTyLeuPheAsnAlaThrLysA 78
   ::::::::::::::::::::::::::::::::::::::::::::::
101 AGCGATTTTATTTCAAAAATGTTGCCATTTTGTATCTCTGAAACATGGAAG 150
   ::::::::::::::::::::::::::::::::::::::::::::::
78 rgArgValPhePheArgAsnIleLysIleLeuIleProAlaThrTrpLys 94
   ::::::::::::::::::::::::::::::::::::::::::::::
151 ACAAGAGTGCATGTATGTGACCAACAACTTGACACCTACAAAATGCTGA 200
   ::::::::::::::::::::::::::::::::::::::::::::::
95 AlaAsnAsnAsn...SerLysIleLysGlnGluSerTyGluLysAlaAs 110
   ::::::::::::::::::::::::::::::::::::::::::::::
201 TGTCTGTTGTTGAGTCTANTCTCCAGGNAATGATGAACCTACACTG 250
   ::::::::::::::::::::::::::::::::::::::::::::::
110 nValIleValThrAspTrpTyGlyAlaHisGlyAspProTyThrL 127
   ::::::::::::::::::::::::::::::::::::::::::::::
251 NGCAGATGGGCAACTGTGGCGAG 273
   ::::::::::::::::::::::::::::::::::::::::::::::
127 euGlnTyArgGlyCysGlyLys 134

```

seq_name: sp_human:Q9Y6N2

seq_documentation_block:

```

ID Q9Y6N2 PRELIMINARY; PRT; 943 AA.
AC Q9Y6N2;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
EX MEDLINE=99292364; PubMed=10362588;
RA Gruber A.D., Schreier K.D., Ji H.-L., Fuller C.M., Pauli B.U.;
RT "Molecular cloning and transmembrane structure of hCLCA2 from human
lung, trachea, and mammary gland.";
RL Am. J. Physiol. 276:C1261-C1270(1999).
RN [2]
RP SEQUENCE FROM N.A.
EX MEDLINE=99364503; PubMed=10437792;
RA Aguel M., Vermaat T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
chloride channel (CaCC) family predominantly expressed in the
digestive tract and trachea.";
RL FESS Lett. 455:295-301(1999).
DR EMBL; AF043977; AAD40367.1; -.
DR EMBL; AF127980; AAD48397.1; -.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 943 AA; 103940 MW; 0F09A09050D2529B CRC64;

```

alignment_scores:

Quality: 223.50 Length: 91
Ratio: 3.020 Gaps: 1
Percent Similarity: 81.319 Percent Identity: 46.154

alignment_block:

US-09-049-696-2 x Q9Y6N2

Align seg 1/1 to: Q9Y6N2 from: 1 to: 943

```

1 GTTGCATCGACCCCAATGTGCCAGAGATGAACACTATTCAACAAT 50
   ::::::::::::::::::::::::::::::::::::::::::::::
45 IleAlaIleAsnProGlnValProGluAsnGlnAsnLeuIleSerAsn11 61
   ::::::::::::::::::::::::::::::::::::::::::::::
51 AAAGGACATGGTGACCCAGCATCTCTGTATCTGTGTTGAAGCTACAGGAA 100
   ::::::::::::::::::::::::::::::::::::::::::::::

```

```
61 eLysGluMetIleThrGluAlaSerPheTyrLeuPheAsnAlaThrLysA 78
101 ACCGATTTTATTCAAAATGTTGCCATTTTCATTCTCAAAACATGGAAG 150
   :::: :::::||||: ||| ||| ||| ||| ||| ||| ||| |||
78 xArgValPhePheArgAsnIleLysIleLeuIleProAlaThrTrpLys 94
151 ACAAGGNTGACTGTGAGACCAAACTTGAGACCTACAAAATGCTGA 200
   ::::: :::: :::: ||| ||| ||| ||| ||| ||| ||| |||
95 AlaAsnAsn...SerLysIleLysGlnGluSerTyrGluLysAlaAs 110
201 TCTTCTGTTGCTGAGTCTANTCTCCAGGNAATGATGAACCTACACTG 250
   :|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
110 nValIleValThrAspTrpTyrGlyAlaHisGlyAspProTyrThrL 127
251 NGCAGATGGCAACTGTGGCGAG 273
   ||| ::||| ||| |||
127 euGlnTyrArgGlyCysGlyLys 134
```


DR GCRDb; GCR_1188; -;
 DR GCRDb; GCR_2514; -;
 DR MIM: 600821; -;
 DR InterPro; IPR000276; GPCR_Rhodopsn.
 DR Pfam; PF00001; 7tm_1; 1.
 DR PRINTS; PRO0237; GPCRHHODPSN.
 DR PRINTS; PRO0752; VASOPRESNVIAR.
 DR PRINTS; PRO0896; VASOPRESNINR.
 DR PROSITE; PS00237; G-PROTEIN RECEPTOR_F1_1; 1.
 DR PROSITE; PS00262; G-PROTEIN RECEPTOR_F1_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.
 FT DOMAIN 1 52 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 53 76 1 (POTENTIAL).
 FT DOMAIN 77 88 2 (POTENTIAL).
 FT TRANSMEM 89 110 3 (POTENTIAL).
 FT DOMAIN 111 125 4 (POTENTIAL).
 FT TRANSMEM 126 147 5 (POTENTIAL).
 FT DOMAIN 148 168 6 (POTENTIAL).
 FT TRANSMEM 169 190 7 (POTENTIAL).
 FT DOMAIN 191 218 8 (POTENTIAL).
 FT TRANSMEM 219 239 9 (POTENTIAL).
 FT DOMAIN 240 293 10 (POTENTIAL).
 FT TRANSMEM 294 313 11 (POTENTIAL).
 FT DOMAIN 314 330 12 (POTENTIAL).
 FT TRANSMEM 332 351 13 (POTENTIAL).
 FT DOMAIN 352 418 14 (POTENTIAL).
 FT CARBOHYD 27 27 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT DISULFID 124 203 BY SIMILARITY.
 SQ SEQUENCE 418 AA; 46799 MW; 477D4260EC826561 CRC64;

alignment_scores:
 Quality: 63.50 Length: 69
 Ratio: 1.984 Gaps: 2
 Percent Similarity: 46.377 Percent Identity: 26.087

alignment_block:
 US-09-049-696-2 x VLAR_HUMAN ..

Align seg 1/1 to: VLAR_HUMAN from: 1 to: 418

2 TTGCAATGACCCCAATGTGCGAAGATGAACACATTCATCAACAATA 51
 :||||| ||| |||||:|||||:|||||:|||||
 360 ValGlnSerPheProCysCysGlnAsnMetLysGluLysPheAsnLys.. 375
 :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
 52 AAGGACATGGTGACCCGCGCATCTCTGTATCTGTGTTGAAGCTACAGGAAA 101
 :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
 376GluAspT 378
 102 GCGATTTTATTTCAAAATGTTGCCATTTTGCATTCCTGAAACATGGAGA 151
 :|||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
 378 hrAspSerMetSerArg..... 383
 152 CAAGGNTGACTGTGAGACCAAACTTGAGACTTGCATCAACAAATGCTGAT 201
 :|||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
 384 ...ArgGlnThrPheTyrSerAsnAsnArgSerProThrAsnSerThrG1 399
 :|||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
 202 GTTCTGG 208
 :|||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
 399 yMetTrp 401

seq_name: SwissProt_39:HMZ1_ARATH

seq_documentation_block:
 ID HMZ1_ARATH STANDARD; PRT; 466 AA.
 AC P42043;

DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE FERROCHELATASE I, CHLOROPLAST/MITOCHONDRIAL PRECURSOR (EC 4.99.1.1)
 DE (PROTOHEME FERRO-LYASE) (HEME SYNTHETASE).
 GN HEM15 OR AT5G26030 OR TLN2.17.

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. LANDSBERG ERECTA; TISSUE=Seedling;
 RX MEDLINE=94230447; PubMed=8175771;
 RA Smith A.G., Santana M.A., Wallace-Cook A.D.M., Roper J.M.,
 RA Labbe-Bois R.;
 RT "Isolation of a cDNA encoding chloroplast ferrochelatase from
 Arabidopsis thaliana by functional complementation of a yeast
 mutant.";
 RL J. Biol. Chem. 269:13405-13413(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. LANDSBERG ERECTA;
 RX MEDLINE=98426386; PubMed=9753778;
 RA Chow K.-S., Singh D.P., Walker A., Smith A.G.;
 RT "Two different genes encode ferrochelatase in Arabidopsis: mapping,
 expression and subcellular targeting of the precursor proteins.";
 RL Plant J. 15:531-541(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=21016721; PubMed=11130714;
 RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
 RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
 RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
 RA Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
 RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
 RA Habermann K., Murray J., Johnson D., Kohlring T., Nelson J.,
 RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
 RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
 RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
 RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,
 RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,
 RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
 RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Marra M.,
 RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
 RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
 RA Entian K.-D., Terry N., Hartley N., Bent E., Johnson S.,
 RA Langham S.-A., McCullagh B., Robben J., Grynoprez B., Zimmermann W.,
 RA Ransperger U., Wedler H., Balke K., Wedler E., Peters S.,
 RA van Staveren M., Dirkse W., Moolman P., Klein Lankhorst R.,
 RA Weitzenegger T., Bothe G., Rose M., Hauf J., Berner S., Hempel S.,
 RA Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiles W.,
 RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,
 RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.F.;
 RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
 thaliana.";
 RL Nature. 408:823-826(2000).
 RN [4]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=97354168; PubMed=9210462;
 RA Roper J.M., Smith A.G.;
 RT "Molecular localisation of ferrochelatase in higher plant
 chloroplasts.";
 RL Eur. J. Biochem. 246:32-37(1997).
 RN [5]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=98010585; PubMed=9346891;
 RA Chow K.-S., Singh D.P., Roper J.M., Smith A.G.;
 RT "A single precursor protein for ferrochelatase-I from Arabidopsis is
 imported in vitro into both chloroplasts and mitochondria.";
 RL J. Biol. Chem. 272:27565-27571(1997).
 CC -1- FUNCTION: CATALYZES THE FERROUS INSERTION INTO PROTOPHYRIN IX.
 CC MAY HAVE A ROLE IN DEALING WITH OXIDATIVE STRESS.
 CC -1- CATALYTIC ACTIVITY: PROTOPHYRIN + Fe(2+) = PROTOHEME + 2 H(+).
 CC -1- PATHWAY: LAST STEP IN PROTOHEME BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (THYLAKOID AND ENVELOPE

[illegible]

```

197 CTGATGTTCTG.....GTTGCTGAGTCTANT 222
   :: :: :: ||| |||
299 roGlnGluValMetIlePheSerAlaHisGlyValProValSerTyr 315
   :: :: :: ||| |||
223 CCTCAGGNAATGATGAACCTACACTGCGCAGATGGCAACTGT 267
   :: :: :: ||| |||
316 ValGluAsnAlaGlyAspProTyrLysAspGlnMetGluGluCys 330
   :: :: :: ||| |||
seq_name: SwissProt_39:LY41_AQUAE

```

```

seq_documentation_block:
ID LY41_AQUAE STANDARD; PRT; 528 AA.
AC O6682;
DT 30-MAY-2000 (Rel. 39, Created)
DE 30-MAY-2000 (Rel. 39, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE LYASE AQ_356 (EC 4.1.3.-).
GN AQ_356.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus";
RL Nature 392:353-358(1998).
CC -1: SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE
CC SYNTHASE FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000685; AAC06637.1; -
DR InterPro: IPR002034; ALPM_homocit_synth.
DR Pfam; PF00682; HMGL-like; 1.
DR PROSITE; PS00815; ALPM_HOMOCIT_SYNTH_1; 1.
DR PROSITE; PS00816; ALPM_HOMOCIT_SYNTH_2; FALSE NEG.
KW Hypothetical protein; Lyase; Complete proteome.
SQ SEQUENCE 528 AA; 59214 MW; 1BC6ABAE3D5D3F34 CRC64;

```

```

alignment_scores:
  Quality: 63.50 Length: 101
  Ratio: 1.176 Gaps: 6
Percent Similarity: 53.465 Percent Identity: 26.733
alignment_block:
US-09-049-696-2 x LY41_AQUAE ..
Align seg 1/1 to: LY41_AQUAE from: 1 to: 528

1 GTTGAATGACCCCAATGTCGACAGATGAACACTCATTCACAAAT 50
   :: ||||| :: ||||| :: ||||| :: ||||| :: |||||
349 ileGluIleAspProLysSerProGluLeuLysLysLeuIleAspLys11 365

51 AAAGGACATGGTGACCCAGGACTCTCTATCTGTTTGAAGCTACAGGA 100
   ||||| :: ||| ||||| :: ||||| :: ||||| :: |||||
365 elysGluLeuGluLysGluGly.....TyrHisPheGluAlaAlaGlu. 379

101 AGCGATTTTATTTCAAAATGTTGCCATTTTTCATTCCTGAACATCGAAG 150
   :: :: :: ||||| :: ||||| :: ||||| :: |||||
380 .....AlaSerLeuGluLeuIleLysArgHisPheGly 391

```

seq_name: SwissProt_39:CLD4_MOUSE

seq_documentation_block:

```

ID CLD4_MOUSE STANDARD; PRT; 210 AA.
AC C35054;
DT 30-MAY-2000 (Rel. 39, Created)
DE 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLAUDIN-4 (CLOSTRIDIUM PERFRINGENS ENTEROTOXIN RECEPTOR) (CPE-
DE RECEPTOR) (CPE-R).
GN CLDN4 OR CPETR1 OR CPER.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97476271; PubMed=9334247;
RA Katsuhira J., Sugiyama H., Inoue N., Horiguchi Y., Matsuda M.,
RA Sugimoto N.;
RT "Clostridium perfringens enterotoxin utilizes two structurally related
RT membrane proteins as functional receptors in vivo.";
RL J. Biol. Chem. 272:26652-26658(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99110921; PubMed=9892664;
RA Motila K., Furuse M., Fujimoto K., Tsukita S.;
RT "Claudin multigene family encoding four-transmembrane domain protein
RT components of tight junction strands.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:511-516(1999).
CC -1: FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1: SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1: SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB000713; BAA22985.1; -
DR EMBL; AF087822; AAD09757.1; -
DR MGD; MGI:1313314; Cldn4.
DR InterPro: IPR001832; Claudin.
DR InterPro; IPR000729; PMP22_Claudin.
DR Pfam; PF00822; PMP22_Claudin; 1.
DR PRINTS; PR01077; CLAUDIN.
DR PROSITE; PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
SQ SEQUENCE 210 AA; 22338 MW; 3B6D571EC71D6564 CRC64;

```

```
CC CC INDICATION: BY HEAT SHOCK:
CC CC -1- SIMILARITY: TO OTHER FUNGAL TPS1 AND TO THE E. COLI ENZYME (GENE OTSA).
CC CC -2- SIMILARITY: CONTAINS A DOMAIN FOUND IN ALL THE PROTEINS OF THE TPS COMPLEX.
CC CC -3- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 475 ONWARD AND IS SHORTER (479 AA) DUE TO A FRAMESHIFT.
CC CC
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC CC
DR DR EMBL; Z29971; CAA82861.1; ALT_FRAME.
DR DR EMBL; AL360054; CAB95998.1; -.
DR DR InterPro: IPR001830; Glyco_transf_20.
DR DR Pfam: PF00982; Glyco_transf_20; 1.
KW K Transferase; Glycosyltransferase; Glycolysis.
FT DOMAIN 1 479 TPS COMPLEX DOMAIN.
FT CONFLICT 277 277 A -> R (IN REF. 1).
SQ SEQUENCE 513 AA; 58493 MW; 9FC247B626CE2B00 CRC64;

alignment_scores:
Quality: 62.50 Length: 84
Ratio: 1.562 Gaps: 5
Percent Similarity: 47.619 Percent Identity: 27.381

alignment_block:
US-09-049-696-2 x TPS1_SCHPO ..
Align seg 1/1 to: TPS1_SCHPO from: 1 to: 513

19 GTGCCAGGAAGATGAA.....ACACTCATTTCAACAATAAGGCATGTGT 62
   : ::::::::::::::::::::
72 lIeProGlNAspGLuYsPrometillelleGlnArgLeuGlnAspGLucy 88
```

```

05 GACCCAGGCGATCTCTGATCTGTTTGAGACCTACAGGAAAGCGGATTTTAAI... 111
   ... : ::::: ||| ::: ||::: ||| |||
88 sSerAlaIleProValPheLeuAspAspGluThrAlaAspArgHisTyrA 105
   ... : ::::: ||| ::: ||::: ||| |||
112 .....TTCAAAAGATGTCGCAATTTTGATTCCT..... 138
   |||::: ||| ::: ||| |||
105 sNGlyPheSerAsn...SerIleLeuTyrProLeuPheHisTyrHisPro 120
   ... : ::::: ||| ::: ||::: ||| |||
139 .....GAAACATGGAGACAAAGGNGTACTATGT 167
   ... : ::::: ||| ::: ||::: ||| |||
121 GlyGluIleAsnPheAspGluGluAsnTrp..... 130
   ... : ::::: ||| ::: ||::: ||| |||
163 GAGACCAAACTTGAGACCTACAAAATGCTGATCTTCTGTTGCTGAGT 217
   ... : ::::: ||| ::: ||::: ||| |||
131 .....GluAlaTyrArgAlaAlaAsnTyrAlaPheAlaGluA 143
   ... : ::::: ||| ::: ||::: ||| |||
218 CT 219
   ..
143 la 143

seq_name: SwissProt_39:DP2S_PYRHO
seq_documentation_block:
ID DP2S_PYRHO STANDARD; PRT; 622 AA.
AC Q57863;
DT *20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA POLYMERASE II SMALL SUBUNIT (EC 2.7.7) (POL II).
GN POLB OR PH0123.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.

```


[8]
RX X-RAY CRYSTALLOGRAPHY (2.85 ANGSTROMS) OF 30-211 IN COMPLEX WITH TNFB.
Banner D.W., D'Arcy A., Janes W., Gentz R., Schoenfeld H.-J.,
RA Broger C., Loetscher H., Lesslauer W.;
RT "Crystal structure of the soluble human 55 kd TNF receptor-human TNF
RL beta complex: implications for TNF receptor activation";
Cell 73:431-445(1993).
[9]
RX X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 41-202.
MEDLINE=97094982; PubMed=8939750;
RA Naimith J.H., Devine T.O., Khono H., Sprang S.R.;
RT "Structures of the extracellular domain of the type I tumor necrosis
RL factor receptor"; J Biol Chem 271:1262(1996).
[10]
CC -!- FUNCTION: RECEPTOR FOR TNF-ALPHA. THE ADAPTOR MOLECULE FADD
CC RECRUITS CASPASE-8 TO THE ACTIVATED RECEPTOR. THE RESULTING
CC AGGREGATE CALLED THE DEATH-INDUCING SIGNALING COMPLEX (DISC)
CC PERFORMS CASPASE-8 PROTEOLYTIC ACTIVATION WHICH INITIATES THE
CC SUBSEQUENT CASCADE OF CASPASES (ASPARTATE-SPECIFIC CYSTEINE
CC PROTEASES) MEDIATING APOPTOSIS. CONTRIBUTES TO THE INDUCTION OF
CC NONCYTOTOXIC TNF EFFECTS INCLUDING ANTI-VIRAL STATE AND ACTIVATION
CC OF THE ACID SPHINGOMYELINASE.
CC -!- SUBUNIT: TNF BINDING TO THE EXTRACELLULAR DOMAIN OF TNFR1 LEADS TO
CC HOMOTRIMERIZATION. ONCE AGGREGATED THE RECEPTORS DEATH DOMAINS
CC PROVIDE A NOVEL MOLECULAR INTERFACE THAT INTERACTS SPECIFICALLY
CC WITH THE DEATH DOMAIN OF TRADD. VARIOUS TRADD-INTERACTING
CC PROTEINS SUCH AS TRAFs, RIP AND POSSIBLY FADD, ARE RECRUITED TO
CC TNFR1 COMPLEX BY THEIR ASSOCIATION WITH TRADD. THIS COMPLEX
CC ACTIVATES AT LEAST TWO DISTINCT SIGNALING CASCADES, APOPTOSIS AND
CC NF-KAPPA B SIGNALING.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- DOMAIN: THE DOMAIN THAT INDUCES A-SMASE IS PROBABLY IDENTICAL TO
CC THE DEATH DOMAIN. THE N-SMASE ACTIVATION DOMAIN (NSD) IS BOTH
CC NECESSARY AND SUFFICIENT FOR ACTIVATION OF N-SMASE.
CC -!- SIMILARITY: CONTAINS A LA-NGFR/TNFR-TYPE CYSTEINE-RICH REGION.
CC -!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD120a entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd120a.htm".

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X55313; CAA39021.1; -
DR EMBL; M33294; AAA03210.1; -
DR EMBL; M58286; AAA36753.1; -
DR EMBL; M63121; AAA36754.1; -
DR EMBL; M75866; AAA61201.1; -
DR EMBL; M75864; AAA61201.1; JOINED.
DR EMBL; M75865; AAA61201.1; JOINED.
DR EMBL; M60275; AAA36756.1; -
DR EMBL; A21522; CAA01558.1; -
DR PIR; A34899; GQHUT1.
DR PIR; A35010; A35010.
DR PIR; S12057; S12057.
DR PIR; A38208; A38208.
DR PDB; 1TNR; 31-JUL-94.
DR PDB; 1NCF; 07-DEC-95.
DR PDB; 1EXT; 11-JAN-97.
DR MIM; 191190; -
DR InterPro; IPR000488; Death.
DR InterPro; IPR001368; TNFR_c6.
DR Pfam; PF00531; death; 1.
DR Pfam; PF00020; TNFR_c6; 4.
DR ProDom; PD000771; TNFR_c6; 1.
DR SMART; SM00005; DEATH; 1.
DR SMART; SM00208; TNFR; 4.
DR PROSITE; PS00652; TNFR_NGFR_1; 3.

DR PROSITE; PS50050; TNFR_NGFR_2; 3.
KW RECEPTOR; Transmembrane; Glycoprotein; Repeat; Signal; Apoptosis;
KW 3D-structure.
FT SIGNAL 1 21
FT CHAIN 22 455 TUMOR NECROSIS FACTOR RECEPTOR 1.
FT CHAIN 41 291 TUMOR NECROSIS FACTOR BINDING PROTEIN 1.
FT DOMAIN 22 211 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 212 234 POTENTIAL.
FT DOMAIN 235 455 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 43 196 4 X TNFR-CYS.
FT REPEAT 43 82 TNFR-CYS 1.
FT REPEAT 83 125 TNFR-CYS 2.
FT REPEAT 126 166 TNFR-CYS 3.
FT REPEAT 167 196 TNFR-CYS 4.
FT DOMAIN 338 348 N-SMASE ACTIVATION DOMAIN (NSD).
FT DOMAIN 356 441 DEATH.
FT DISULFID 44 58
FT DISULFID 59 72
FT DISULFID 62 81
FT DISULFID 84 99
FT DISULFID 102 117
FT DISULFID 105 125
FT DISULFID 127 143
FT DISULFID 146 158
FT DISULFID 149 166
FT DISULFID 168 179
FT DISULFID 182 191
FT DISULFID 185 195
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 412 412 MISSING (IN REF. 4).
FT CONFLICT 443 446 GPAA -> APP (IN REF. 4).
SQ SEQUENCE 455 AA; 50494 MW; 4CEFA96D03B8225 CRC64;

alignment_scores:
Quality: 62.00 Length: 60
Ratio: 2.067 Gaps: 3
Percent Similarity: 50.000 Percent Identity: 31.667

alignment_block:
US-09-049-696-2/rev x TNFR1_HUMAN
Align seg 1/1 to: TNFR1_HUMAN from: 1 to: 455

177 TTTTGGTCTCAGATAGTCACATGCTTTCATGCTTTCAGGAATCAAAA 128
135 TyrTrpSerGluAsnLeuPheGlnCysPheAsnCysSerLeuCysLeuAs 151
127 TGGCAACATTTTGGAAATAAAATCGCTTTCCTCTAGCTTCAACAGATAC 78
151 nGlyThrVal.....HisLeuSerCysGlnGluLysGlnAsn 164
77 AGAGATCCCTGGGTGCACCATGCTCTTTATTGTTG...AATGAGTGTTC 31
164 hrValCysThrCysHisAlaGlyPhePheLeuArgGluAsnGluCys... 179
130 ATCTTCTGGCACATTTGGGGTGCATTTGCAAC 1
180ValSerCysSer 183

This Page Blank (uspto)

OM of: US-09-049-696-2 to: PIR_68:* out_format : pfs
Date: Mar 30, 2002 2:26 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODL=frame+ntp.model -DEV=xlp
-O/cgn2_1/JUSPTO_spool/US09049696/runat_28032002_145238_2031/app_query.fasta_1.12579
-DB=PIR_68 -QFMT=Fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09049696_@CEN1_1_367 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-049-696-2
Query length: 273
Database: PIR_68:*
Database sequences: 219241
Database length: 76174552
Search time (sec): 470.790000

score_list:	Strtd Orig	ZScore	Escore	Len	Documentation
Sequence					
pir2:JG0168	+ 371.00	807.90	5.7e-38	913	! gob-5 protein - mouse
pir2:T02205	+ 288.00	621.70	1.4e-27	905	! Lu-ECAM-1 protein - bovine
pir2:E72419	+ 81.00	157.77	0.1016	858	! flagellar hook-associated prote
pir2:E71451	+ 75.50	147.76	0.4871	630	! hypothetical protein PH0268 - B
pir2:T10997	- 68.00	141.46	3.95	174	! NADH dehydrogenase (ubiquinone)
pir2:T09621	+ 67.50	126.83	4.96	906	! probable disease resistance pro
pir2:T39698	+ 66.50	123.86	6.29	330	! protein tyrosine phosphatase -
pir2:T04373	+ 66.50	129.72	6.41	484	! probable ferredoxin-like (EC 4.9
pir2:T05736	+ 66.50	129.72	6.41	484	! probable ferredoxin-like (EC 4.9
pir2:T05199	+ 66.50	127.70	6.49	619	! DNA-directed DNA polymerase (EC
pir2:T03896	- 66.50	125.55	6.58	805	! hypothetical protein R09E12 - 6
pir2:H75174	+ 66.00	134.04	7.16	249	! ATP-binding protein pmp (pmp)
pir2:A28568	+ 66.00	128.82	7.39	471	! pneumolysin - Streptococcus pne
pir2:T01312	+ 65.50	134.47	8.19	206	! hypothetical protein T14P8.2 -
pir2:JX0369	+ 65.50	120.44	8.93	1142	! collagen alpha 1(XIX) chain pr
pir2:T28048	+ 65.50	118.26	9.05	1491	! hypothetical protein ZK956.9 -
pir2:T52137	+ 65.00	124.34	10.00	619	! ATP-dependent DEAD box RNA heli
pir2:S45768	+ 65.00	117.01	10.46	1513	! mitotic spindle protein STU1 -
pir2:T43933	+ 64.50	123.29	11.54	613	! DNA-directed DNA polymerase (EC
pir2:A53046	+ 63.50	124.19	15.09	418	! vasopressin receptor via - huma
pir2:C64654	+ 63.50	123.80	15.13	438	! conserved hypothetical secreted
pir2:H71943	+ 63.50	123.80	15.13	438	! hypothetical protein jhp0350 -
pir2:A54125	+ 63.50	123.30	15.17	466	! ferredoxin-like (EC 4.99.1.1) pr
pir2:T10246	+ 63.50	122.49	15.25	514	! ferredoxin-like (EC 4.99.1.1) -
pir2:F70331	+ 63.50	122.27	15.27	528	! 2-isopropylmalate synthase - Ag
pir2:T69155	+ 63.00	124.43	17.28	354	! hypothetical protein MTH428 - M
pir2:T15628	+ 63.00	114.66	18.36	1166	! hypothetical protein C25H3.9 -
pir2:T46564	+ 62.50	120.83	20.26	479	! alpha,alpha-trehalose-phosphate
pir1:A71233	+ 62.50	118.69	20.53	622	! probable DNA-directed DNA polym
pir2:A84685	+ 62.50	116.13	20.86	850	! probable RNA methyltransferase
pir2:T18451	+ 62.50	115.56	20.93	911	! hypothetical protein C0505w - m
pir2:T33783	+ 62.50	109.97	21.67	1802	! hypothetical protein Y39D8C.1
pir2:T83910	+ 62.00	126.98	22.37	137	! hypothetical protein BH2088.1
pir2:T71867	- 62.00	122.61	22.98	335	! hypothetical protein jhp0966 -
pir2:A84680	- 62.00	121.41	23.15	389	! hypothetical protein jhp0966 -
pir2:T84879	- 62.00	121.26	23.18	396	! hypothetical protein At228040
pir1:GQHUT1	- 62.00	120.12	23.34	455	! hypothetical protein At228010
pir2:H96650	- 62.00	111.47	24.62	1309	! protein T3P18.3 (imported) - A
pir2:C86438	- 62.00	110.83	24.72	1415	! hypothetical protein AAD21687 -
pir2:T29415	- 61.50	121.22	26.58	347	! hypothetical protein F21P8.9 -
pir2:JC4302	- 61.50	118.90	26.97	461	! tumor necrosis factor receptor

pir2:T29642 - 61.50 118.07 27.11 510 ! hypothetical protein F49E8.7
pir2:F83681 + 61.50 117.14 27.26 571 ! urease alpha subunit urec [i
pir1:VCPV2M + 61.50 115.29 27.58 716 ! coat protein VP1 - minute vi
pir2:T23810 + 61.50 113.78 27.84 861 ! hypothetical protein M28.9 -
seq_name: pir2:JG0168
seq_documentation_block:
C:Species: Mus musculus (house mouse)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
C:Accession: JG0168
R:Komiyama, T.; Tanigawa, Y.; Hirohashi, S.
Biochem Biophys Res Commun. 255, 347-351, 1999
A:Title: Cloning and identification of the gene gob-5, which is expressed in intestin
A:Reference number: JG0168; MUID:99160866
A:Accession: JG0168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-913 <KOM>
A:Cross-references: DDBJ:AB016592
alignment_scores:
Quality: 371.00 Length: 91
Ratio: 4.417 Gaps: 0
Percent Similarity: 92.308 Percent Identity: 76.923
alignment_block:
US-09-049-696-2 x JG0168
Align seg 1/1 to: JG0168 from: 1 to: 913
1 GTTGCATCGACCCCAATGTCGAGAGATGAAACACTATTCAACAAAT 50
37 IleAlaIleAspHisaspValProGluaspGluAlaLeuileGlnHisII 53
51 AAAGGACATGGTGACCCAGCATCTCTGTATCTGTGTGTAAGCTACAGGAA 100
53 eLysaspMetValThrGlnAlaSerProtyrLeuPheGluAlaThrGlyL 70
101 AGCGATTTTATTTCAAAATGTTGCCATTTTGTATCTGCTGAACATGAAG 150
70 ysargPheTyrPheLysAsnValAlaLeuileProGluSerTrpLys 86
151 ACAAGGNTGACATGATGACACCAAACTTGCAGACCTACACCAATGCTGA 200
87 AlaLysProGluTyThrArgProLysLeuGluThrPheLysAsnAlaAs 103
201 TGTCTGTTGCTGAGTCTANTCTCCAGGNAATGATGACACCTACACTG 250
103 pValLeuValSerThrThrSerProLeuGlyAsnAspGluProtyrThrG 120
251 NGCAGATGGCAACTGTGGCGAG 273
120 LuHisileGlyAlaCysGlyGlu 127
seq_name: pir2:T02205
seq_documentation_block:
Lu-ECAM-1 protein - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 01-Dec-2000
C:Accession: T02152; T02171
R:Elble, R.C.; Widom, J.; Gruber, A.D.; Abdel-Ghany, M.; Levine, R.; Goodwin, A.; Pau
submitted to the EMBL Data Library, April 1997
A:Description: Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial
A:Reference number: Z14590
A:Accession: T02205
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-905 <ELB>
A:Cross-references: EMBL:AF001261; NID:g2623762; PIDN:AB86529.1; PID:g2623763

A: Experimental source: lung
A: Accession: T02152
A: Status: preliminary; translated from GB/EMBL/DDBJ
A: Molecule type: mRNA
A: Residues: 1-793, 'SGSFP', 'M', '862', 'RE', '865-866', 'Q', '868', 'AKVLELO', '876', 'QHQ', '880', 'FQ' <EL3>
A: Cross-references: EMBL:AF001263; NID:g2623766; PIDN:AAB86531.1; PID:g2623767
A: Experimental source: lung
A: Accession: T02171
A: Status: preliminary; translated from GB/EMBL/DDBJ
A: Molecule type: mRNA
A: Residues: 1-792, 'ES' <EL3>
A: Cross-references: EMBL:AF001262; NID:g2623764; PIDN:AAB86530.1; PID:g2623765
A: Experimental source: lung

alignment_scores:
Quality: 288.00 Length: 91
Ratio: 3.840 Gaps: 0
Percent Similarity: 82.418 Percent Identity: 59.341

alignment_block:
US-09-049-696-2 x T02205 ..

Align seg 1/1 to: T02205 from: 1 to: 905

1 GTTGCAATCGACCCCAATGTGCCAGAAATGACACACTCATTTCAACAAT 50
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
36 lleAlalleasnProSerValProgluaspglulysLeuilleGluasnI 52

51 AAAGGACATGGTGCACCCAGGAGCTCTCTATCTGTTTGAAGCTACAGGAA 100
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
52 elysGluMetValThrGluAlaSerThrTyrlouPheHisAlatThrLysA 69

101 ACGCATTATTTAAAAAATGTTGCCATTTTTCCTGTAACATGGAAG 150
::::|||:::|||||:::|||||:::|||||:::|||||:::|||||
69 rgargValTyrrPheargAsnvalSerilleulleleuproMetThrTrplys 85

151 ACAAGAGNTGACTATGTGAGACCAAAATCTGAGACTACAAAATGCTCA 200
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
86 SerLysSerGluTyrrPheilleProlysGlnGluSerTyrrAspGlnalaas 102

201 TGTTCTGGTTGCTGACTCTANTCTCCAGGNAATGATGAACCTCACACTG 250
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
102 pValilleValAlaAsnProTyrrLeuLysTyrrGlyAspproTyrrThRL 119

251 NGCAGATGGGCAACTGTGCGCAG 273
||| |||:::|||||
119 euGlnTyrrGlyArgCysGlyGlu 126

seq_name: pir2:E72419

seq_documentation_block:
flagellar hook-associated protein 1 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: E72419
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Winn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.
C.M.

Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316
A:Accession: E72419
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-858 <ARN>
A:Cross-references: GB:AF001695; NID:g4980569; PIDN:AAD35177.1; PID:g498057
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0083

alignment_scores:


```

F:292-1123/Region: interrupted helical
F:1124-1142/Domain: carboxyl-terminal nonhelical #status predicted <NH2>

alignment_scores:
  Quality: 65.50      Length: 98
  Ratio: 1.637       Gaps: 3
  Percent Similarity: 40.816  Percent Identity: 23.469

alignment_block:
  US-09-049-696-2 x JX0369 ..

Align seg 1/1 to: JX0369 from: 1 to: 1142

113  CCCAATGTGCCAGAAGATGAACACTCATTCAACAATAAAGCACATGCT 62
114  |||||:||||| ||| :|||:|||||:||||| ||| :|||:
1006 ProGlyIleProAlaAspAlaValSerPheGluGluIleLysLysTyrI1 1022
163  GACCCAGGCATCTGTGTATCTGTTGAAGCTCAGGAAAGCCGATTATT 112
164  ||||| ||| :|||:||||| |||
1022 eAsGInGluValLeuArgIlePheGluGluArgMetAlaValPheLeuS 1039
113  TCAAAATGTGGCATTTTTCCTGAAACATGGAGACAAAGGNTGAC 162
164  ||| :|||:||||| ||| :|||:
1039 erGIn.....LeuLysLeuProAlaAlaMetLeuAlaGlnAla 1052
163  TATGTGAGACCAAACTTGGACACCTCACAAAATGCTGTGTTCTGTTCC 212
164  ||| |||||
1053 TyrGlyArgPro..... 1056
213  TGAGTCTANTCCTCCAGGNAATGAT..... 237
1057 .....GlyProGlyLysaspGlyLeuProGlyProGlyAspP 1071
238 .....GAACCCACACGTNGCAGATGGCAACTGTGGCGAG 273
1071 roGlyProGInGlyTvrArgGlyGlnLysGlyGluArgGlyGlu 1085

```

This Page Blank (uspto)


```
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-09-446-504-5

alignment_scores:
    Quality: 64.50      Length: 55
    Ratio: 1.697        Gaps: 2
    Percent Similarity: 69.091   Percent Identity: 23.636

alignment_block:
US-09-049-696-2 x US-09-446-504-5 ..
Align seg 1/1 to: US-09-446-504-5 from: 1 to: 613

1 GTTCCAATCGACCCCATGTGCCAGAAGATGAACAACATCATTCAACAATAAT 50
   ||| ||| ||||| :||||||| :||||:|||||
531 ValProIleAlaProAsp...ProGluAspLeuValIleGluGluVa 546
   : |||:||| :|||: :|||:|||||:|||||:
546 lProAspValHisMetglyHisValHisValTyAspAlaVal.... 561

101 AGCGATTTTATTCAAAANGTTGCCATTTTGCTATCTGTATCTGTGAAGCTACAGGAA 100
   : |||:||| :|||: :|||:|||||:|||||:
562 .....ValTyArgGlyValGlnLeuValAsnSerAlaThrTrpGln 575

151 ACAAAGNGTACTAT 165
   :|||:|||||:
576 AlaGlnThrGluPhe 580

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pap:US-08-050-319B-52

seq_documentation_block:
; Sequence 52, Application US/08050319B
; Patent No. 5633145
; GENERAL INFORMATION:
; APPLICANT: M.Feldmann, P.W.Gray.
; APPLICANT: M.J.C.Turner,F.M.Brennan
; TITLE OF INVENTION: Modified human TNFalpha (Tumor
; TITLE OF INVENTION: Necrosis Factor alpha) Receptor
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robbins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050.319B
; FILING DATE: 10-May-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Robbins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5150-0030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 153 amino acids
; TYPE: amino acid
```

```
; MOLECULE TYPE: protein
US-08-463-982-52

alignment_scores:
  Quality: 62.00      Length: 60
  Ratio: 2.067        Gaps: 3
Percent Similarity: 50.000 Percent Identity: 31.667

alignment_block:
US-09-049-696-2/rev x US-08-465-982-52 ..
Align seg 1/1 to: US-08-465-982-52 from: 1 to: 153

177 TTTTGGTCTCACATAGTCANCCCTTTGCTTCCATGTTTCCAGGATCAAAA 128
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
89 TyrTrpSerGluAsnLeuPheGlnCysPheAsnCysSerLeuCysLeuAs 105
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
127 TGGCACATTTTGAATAAATCGCTTCTCTAGCTTCAACAGATAC 78
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
105 nGlyThrVal.....HisLeuSerCysGlnGluGlnAsnT 118
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
77 AGAGATCGCTGGGTCCACATGCTCTTATTGTTG...AATGAGTGTTTC 31
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
118 hrValCysThrCysHisAlaGlyPhePheLeuArgGluAsnGluCys... 133
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
30 ATCTTCTGGCACATTTGGGTGCGATTGCAAC 1
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
134 .....ValSerCysSer 137
```

```
seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-219-237B-4

seq_documentation_block:
; Sequence 4, Application US/08219237B
; Patent No. 5874546
; GENERAL INFORMATION:
; APPLICANT: ITOH, Naoto
; APPLICANT: YONEHARA, Shin
; APPLICANT: NAGATA, Shigekazu
; TITLE OF INVENTION: DNA Coding for Human Cell Surface Antigen
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James W. Hellwege
; STREET: P.O. Box 2266 Eads Station
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/219,237B
; FILING DATE: 28-MAR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,129
; FILING DATE: 22-APR-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: James W. Hellwege
; REGISTRATION NUMBER: 28,808
; REFERENCE/DOCKET NUMBER: 516762
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 153 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-219-237B-4
```

```
alignment_scores:
  Quality: 62.00      Length: 60
  Ratio: 2.067        Gaps: 3
Percent Similarity: 50.000 Percent Identity: 31.667

alignment_block:
US-09-049-696-2/rev x US-08-219-237B-4 ..
Align seg 1/1 to: US-08-219-237B-4 from: 1 to: 153

177 TTTTGGTCTCACATAGTCANCCCTTTGCTTCCATGTTTCCAGGATCAAAA 128
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
93 TyrTrpSerGluAsnLeuPheGlnCysPheAsnCysSerLeuCysLeuAs 109
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
127 TGGCACATTTTGAATAAATCGCTTCTCTAGCTTCAACAGATAC 78
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
109 nGlyThrVal.....HisLeuSerCysGlnGluGlnAsnT 122
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
77 AGAGATCGCTGGGTCCACATGCTCTTATTGTTG...AATGAGTGTTTC 31
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
122 hrValCysThrCysHisAlaGlyPhePheLeuArgGluAsnGluCys... 137
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
30 ATCTTCTGGCACATTTGGGTGCGATTGCAAC 1
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
138 .....ValSerCysSer 141

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-477-347-12

seq_documentation_block:
; Sequence 12, Application US/08477347
; Patent No. 6232446
; GENERAL INFORMATION:
; APPLICANT: WALLACH, David
; APPLICANT: BIGDA, Jacek
; APPLICANT: BELETSKY, Igor
; APPLICANT: METT, Igor
; TITLE OF INVENTION: TNF LIGANDS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,347
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/115,685
; FILING DATE:
; APPLICATION DATA:
; APPLICATION NUMBER: IL 106271
; FILING DATE: 08-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, G. Kevin
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: WALLACH=10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 153 amino acids
; TYPE: amino acid
```

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-477-347-12

alignment_scores:
Quality: 62.00 Length: 60
Ratio: 2.067 Gaps: 3
Percent Similarity: 50.000 Percent Identity: 31.667

alignment_block:
US-09-049-696-2/rev x US-08-477-347-12 ..

Align seg 1/1 to: US-08-477-347-12 from: 1 to: 153

```
177 TTTTGGCTCAGATAGTCATCAGCTTTGCTTCCATGTTTCAGGAATCAAAA 128
    :|||||: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
93 TyrTrpSerGluAsnLeuPheGlnCysPheAsnCysSerLeuCysLeuAs 109
127 TGGCAACATTTTGAATAAATCGCTTCCGTAGCTTCAACACAGATAC 78
    :|||||: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
109 nGlyThrVal.....HisLeuSerCysGlnGluLysGlnAsnT 122
77 AGAGATGCTGGTGCACCATGCTTTATTGTTG...AATGAGTGTTC 31
    :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
122 hrValCysThrCysHisAlaGlyPhePheLeuArgGluAsnGluCys... 137
30 ATCTTCTGGCACATTTGGGTGCGATTCGAAC 1
    :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
138 .....ValSerCysSer 141
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-476-862-3

seq_documentation_block:
Sequence 3, Application US/08476862
Patent No. 6262239

GENERAL INFORMATION:
APPLICANT: WALLACH, David
APPLICANT: BIGDA, Jacek
APPLICANT: BELETSKY, Igor
APPLICANT: METT, Igor
TITLE OF INVENTION: TNF INHIBITORS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NETMARK
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476,862
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 107267
FILING DATE: 12-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 94039
FILING DATE: 06-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 91229
FILING DATE: 06-AUG-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 90339
FILING DATE: 18-MAY-1989

ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: WALLACH-12A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 153 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-476-862-3

alignment_scores:
Quality: 62.00 Length: 60
Ratio: 2.067 Gaps: 3
Percent Similarity: 50.000 Percent Identity: 31.667

alignment_block:

US-09-049-696-2/rev x US-08-476-862-3 ..

Align seg 1/1 to: US-08-476-862-3 from: 1 to: 153

```
177 TTTTGGCTCAGATAGTCATCAGCTTTGCTTCCATGTTTCAGGAATCAAAA 128
    :|||||: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
93 TyrTrpSerGluAsnLeuPheGlnCysPheAsnCysSerLeuCysLeuAs 109
127 TGGCAACATTTTGAATAAATCGCTTCCGTAGCTTCAACACAGATAC 78
    :|||||: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
109 nGlyThrVal.....HisLeuSerCysGlnGluLysGlnAsnT 122
77 AGAGATGCTGGTGCACCATGCTTTATTGTTG...AATGAGTGTTC 31
    :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
122 hrValCysThrCysHisAlaGlyPhePheLeuArgGluAsnGluCys... 137
30 ATCTTCTGGCACATTTGGGTGCGATTCGAAC 1
    :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
138 .....ValSerCysSer 141
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-468-560C-4

seq_documentation_block:

Sequence 4, Application US/08468560C
Patent No. 6270998

GENERAL INFORMATION:

APPLICANT: NAGATA, Shigekazu

APPLICANT: ITOH, Naoto

APPLICANT: YONEHARA, Shin

TITLE OF INVENTION: DNA CODING FOR HUMAN CELL SURFACE

TITLE OF INVENTION: ANTIGEN

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP.

STREET: P.O. BOX 747

CITY: FALLS CHURCH

STATE: VA

COUNTRY: USA

ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/468,560C

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: MURPHY JR., GERALD M.

```

; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 20-4393P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 153 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-468-560C-4

```

```

alignment_scores:
  Quality: 62.00      Length: 60
  Ratio: 2.067        Gaps: 3
  Percent Similarity: 50.000  Percent Identity: 31.667

```

```
alignment_block:
```

```
US-09-049-696-2/rev x US-08-468-560C-4 ..
```

```
Align seg 1/1 to: US-08-468-560C-4 from: 1 to: 153
```

```

177 TTTTGGTCTCACAATAGTCANCCCTTTGCTTCCATGTTTCAGGAATCAAAA 128
   :::::::::::   ::   :::::   |||
93 TyrTrpSerGluAsnLeuPheGlnCysPheAsnCysSerLeuCysLeuAs 109
127 TGGCACAATTTTGAATAAAATCGCTTCTGTAGCTTCAAAACAGATAC 78
   ||:::||||:   ::   :::::   |||
109 nGlyThrVal.....HisLeuSerCysGlnGluLysGlnAsnT 122
77 AGAGATCGCTGGTGCCACATGCTCTTTATTGTTG...AATGAGTGTTTC 31
   |||   |||   :::::   |||
122 hrValCysThrCysHisAlaGlyPhePheLeuArgGluAsnGluCys... 137
30 ATCTTCTGGCACATTTGGGGTCGATTGCAAC 1
   ||:::||||:
138 .....ValSerCysSer 141

```

```
seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-050-319B-50
```

```

seq_documentation_block:
; Sequence 50, Application US/08050319B
; Patent No. 5633145
; GENERAL INFORMATION:
; APPLICANT: M.Feldmann, P.W. Gray,
; APPLICANT: M.J.C. Turner, F.M. Brennan
; TITLE OF INVENTION: Modified human TNFalpha (Tumor
; TITLE OF INVENTION: Necrosis Factor alpha) Receptor
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050.319B
; FILING DATE: 10-May-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Robbins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5150-0030
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 157 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-050-319B-50

```

```

alignment_scores:
  Quality: 62.00      Length: 60
  Ratio: 2.067        Gaps: 3
  Percent Similarity: 50.000  Percent Identity: 31.667

```

```
alignment_block:
```

```
US-09-049-696-2/rev x US-08-050-319B-50 ..
```

```
Align seg 1/1 to: US-08-050-319B-50 from: 1 to: 157
```

```

177 TTTTGGTCTCACAATAGTCANCCCTTTGCTTCCATGTTTCAGGAATCAAAA 128
   :::::::::::   ::   :::::   |||
93 TyrTrpSerGluAsnLeuPheGlnCysPheAsnCysSerLeuCysLeuAs 109
127 TGGCACAATTTTGAATAAAATCGCTTCTGTAGCTTCAAAACAGATAC 78
   ||:::||||:   ::   :::::   |||
109 nGlyThrVal.....HisLeuSerCysGlnGluLysGlnAsnT 122
77 AGAGATCGCTGGTGCCACATGCTCTTTATTGTTG...AATGAGTGTTTC 31
   |||   |||   :::::   |||
122 hrValCysThrCysHisAlaGlyPhePheLeuArgGluAsnGluCys... 137
30 ATCTTCTGGCACATTTGGGGTCGATTGCAAC 1
   ||:::||||:
133 .....ValSerCysSer 141

```

```
seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-465-982-50
```

```

seq_documentation_block:
; Sequence 50, Application US/08465982
; Patent No. 5863786
; GENERAL INFORMATION:
; APPLICANT: M.Feldmann, P.W. Gray,
; APPLICANT: M.J.C. Turner, F.M. Brennan
; TITLE OF INVENTION: Modified human TNFalpha (Tumor
; TITLE OF INVENTION: Necrosis Factor alpha) Receptor
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,982
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,319
; FILING DATE: 10-May-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Robbins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5150-0030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999

```

us-09-049-696-2.ra1

Tue Apr 2 09:39:56 2002

```
TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 157 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-465-982-50

alignment_scores:
  Quality: 62.00      Length: 60
  Ratio: 2.067       Gaps: 3
  Percent Similarity: 50.000      Percent Identity: 31.667

alignment_block:
US-09-049-696-2/rev x US-08-465-982-50 ..
Align seg 1/1 to: US-08-465-982-50 from: 1 to: 157

177 TTTTGTCTCACATAGTCANCCCTTTGTCTTCATGCTTTCAGGAATCAAAA 128
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
93 TyrTrpSerGluAsnLeuPheGlnCysPheAsnCysSerLeuCysLeuAs 109
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
127 TGCAACATTTTGAATAAAATCGCTTCCTAGCTTCAAAACAGATAC 78
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
109 nGlyThrVal.....HisLeuSerCysGlnGluLysGlnAsnT 122
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
77 AGAGATCGCTGGTGACCATGCTCTTTATTGTTG...AATGAGTGTTC 31
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
122 hrValCysThrCysHisAlaGlyPhePheLeuArgGluAsnGluCys... 137
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
30 ATCTTGGCACATTTGGGTGCTGATTCGAAC 1
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
138 .....ValSerCysSer 141

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pap:US-09-326-394-2

seq_documentation_block:
; Sequence 2, Application US/09326394
; Patent No. 6306820
; GENERAL INFORMATION:
; APPLICANT: Bendele, Alison M.
; APPLICANT: Sennello, Regina M.
; APPLICANT: Edwards, Carl K.
; TITLE OF INVENTION: COMBINATION THERAPY USING A TNF BINDING
; TITLE OF INVENTION: PROTEIN FOR TREATING TNF-MEDIATED DISEASES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 DeHavilland Drive
; CITY: Thousand Oaks
; STATE: CA
; COUNTRY: US
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/326,394
; FILING DATE: 08-DEC-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,587
; FILING DATE: 06-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/036,355
; FILING DATE: 23-JAN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,315
; FILING DATE: 07-FEB-1997

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/052,023
; FILING DATE: 09-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Zindrick, Thomas K.
; REGISTRATION NUMBER: 32,185
; REFERENCE/DOCKET NUMBER: A-430D
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-326-394-2

alignment_scores:
  Quality: 62.00      Length: 60
  Ratio: 2.067       Gaps: 3
  Percent Similarity: 50.000      Percent Identity: 31.667

alignment_block:
US-09-049-696-2/rev x US-09-326-394-2 ..
Align seg 1/1 to: US-09-326-394-2 from: 1 to: 161

177 TTTTGTCTCACATAGTCANCCCTTTGTCTTCATGCTTTCAGGAATCAAAA 128
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
95 TyrTrpSerGluAsnLeuPheGlnCysPheAsnCysSerLeuCysLeuAs 111
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
127 TGCAACATTTTGAATAAAATCGCTTCCTAGCTTCAAAACAGATAC 78
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
111 nGlyThrVal.....HisLeuSerCysGlnGluLysGlnAsnT 124
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
77 AGAGATCGCTGGTGACCATGCTCTTTATTGTTG...AATGAGTGTTC 31
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
124 hrValCysThrCysHisAlaGlyPhePheLeuArgGluAsnGluCys... 139
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
30 ATCTTGGCACATTTGGGTGCTGATTCGAAC 1
   :|||||: : : : : : : : : : : : : : : : : : : : : : : :
140 .....ValSerCysSer 143

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pap:US-08-050-319B-48

seq_documentation_block:
; Sequence 48, Application US/08050319B
; Patent No. 5633145
; GENERAL INFORMATION:
; APPLICANT: M. Feldmann, P.W. Gray.
; APPLICANT: M.J.C. Turner, F.M. Brennan
; TITLE OF INVENTION: Modified human TNFalpha (Tumor
; TITLE OF INVENTION: Necrosis Factor alpha) Receptor
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robbins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,319B
; FILING DATE: 10-May-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Robbins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5150-0030
```


TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 617-8999
 TELEFAX: (415) 327-3231
 INFORMATION FOR SEQ ID NO: 48:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 199 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-050-319B-48

alignment_scores:
 Quality: 62.00 Length: 60
 Ratio: 2.067 Gaps: 3
 Percent Similarity: 50.000 Percent Identity: 31.667

alignment_block:
 US-09-049-696-2/rev x US-08-050-319B-48

Align seg 1/1 to: US-08-050-319B-48 from: 1 to: 199

```

177 TTTTGGTCTCACATAGTCANCCCTTTGTCTTCCATGTTTCAGGAATCAAAA 128
    :::|||||:::   :::   |||:::   |||
135 TyrTrpSerGluAsnLeuPheGlnCysPheAsnCysSerLeuCysLeuAs 151
127 TGGCAACATTTTGAATAAAATCGCTTTCTCTAGCTTCAAAACAGATAC 78
    |||:::   |||:::   |||:::   |||
151 nGlyThrVal.....HisLeuSerCysGlnGluLysGlnAsn 164
77 AGAGATCCCTGGGTCACCATGCTCTTTATTGTTG....AATGAGTGTTC 31
    |||   |||   :::   |||   |||
164 hrValCysThrCysHisAlaGlyPhePheLeuArgGluAsnGluCys... 179
30 ATCTTCTGGCACATTTGGGTCGATTCGAAC 1
    |||:::   |||:::
180 .....ValSerCysSer 183
  
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-465-982-48

seq_documentation_block:
 ; Sequence 48, Application US/08465982
 ; Patent No. 5863786
 ; GENERAL INFORMATION:
 ; APPLICANT: M.Feldmann, P.W. Gray,
 ; APPLICANT: M.J.C. Turner, F.M Brennan
 ; TITLE OF INVENTION: Modified human TNFalpha (Tumor
 ; TITLE OF INVENTION: Necrosis Factor alpha) Receptor
 ; NUMBER OF SEQUENCES: 57
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Reed & Robbins
 ; STREET: 635 Bryant Street
 ; CITY: Palo Alto
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94301
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/465,982
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/050,319
 ; FILING DATE: 10-May-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Robbins, Roberta L.
 ; REGISTRATION NUMBER: 33,208
 ; REFERENCE/DOCKET NUMBER: 5150-0030
 ; TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 617-8999
 TELEFAX: (415) 327-3231
 INFORMATION FOR SEQ ID NO: 48:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 199 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-465-982-48

alignment_scores:
 Quality: 62.00 Length: 60
 Ratio: 2.067 Gaps: 3
 Percent Similarity: 50.000 Percent Identity: 31.667

alignment_block:
 US-09-049-696-2/rev x US-08-465-982-48

Align seg 1/1 to: US-08-465-982-48 from: 1 to: 199

```

177 TTTTGGTCTCACATAGTCANCCCTTTGTCTTCCATGTTTCAGGAATCAAAA 128
    :::|||||:::   :::   |||:::   |||
135 TyrTrpSerGluAsnLeuPheGlnCysPheAsnCysSerLeuCysLeuAs 151
127 TGGCAACATTTTGAATAAAATCGCTTTCTCTAGCTTCAAAACAGATAC 78
    |||:::   |||:::   |||:::   |||
151 nGlyThrVal.....HisLeuSerCysGlnGluLysGlnAsn 164
77 AGAGATCCCTGGGTCACCATGCTCTTTATTGTTG....AATGAGTGTTC 31
    |||   |||   :::   |||   |||
164 hrValCysThrCysHisAlaGlyPhePheLeuArgGluAsnGluCys... 179
30 ATCTTCTGGCACATTTGGGTCGATTCGAAC 1
    |||:::   |||:::
180 .....ValSerCysSer 183
  
```

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-974-022-46

seq_documentation_block:
 ; Sequence 46, Application US/08974022
 ; Patent No. 6015938
 ; GENERAL INFORMATION:
 ; APPLICANT: Boyle, William J.
 ; APPLICANT: Lacey, David L.
 ; APPLICANT: Calzone, Frank J.
 ; APPLICANT: Chang, Ming-Shi
 ; TITLE OF INVENTION: OSTEOPROTEGERIN
 ; NUMBER OF SEQUENCES: 53
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Amgen Inc.
 ; STREET: 1840 Dehavilland Drive
 ; CITY: Thousand Oaks
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 91320-1789
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/974,022
 ; FILING DATE: 12-DEC-1995
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/577,788
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Winter, Robert B.
 ; REFERENCE/DOCKET NUMBER: A-378
 ; INFORMATION FOR SEQ ID NO: 46:
 ; SEQUENCE CHARACTERISTICS:

Tue Apr 2 09:39:56 2002

us-09-049-696-2.ra1

```

; LENGTH: 280 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-974-022-46

alignment_scores:
  Quality: 62.00      Length: 60
  Ratio: 2.067        Gaps: 3
  Percent Similarity: 50.000  Percent Identity: 31.667

alignment_block:
  US-09-049-696-2/rev x US-08-974-022-46 ..
  Align seg 1/1 to: US-08-974-022-46 from: 1 to: 280

177 TTTTGGTCTCACATAGTCANCCCTTTGCTTCATGTTTCAGGAATCAAAA 128
    :::::::::::  ::  :::::::::::  ||
135 TyrTrpSerGluAsnLeuPheGlnCysPheAsnCysSerLeuCysLeuAs 151
    :::::::::::  ::  :::::::::::  ||
127 TGGCAACATTTTGAATAAATCGCTTTCCTGTAGCTTCAACAGATAC 78
    :::::::::::  ::  :::::::::::  ||
151 nGlyThrVal.....HisLeuSerCysGlnGluLysGlnAsnT 164
    :::::::::::  ::  :::::::::::  ||
77 AGAGATGCCCTGGGTCCACCATGCTTATTTGTTG...AATGAGTGTTC 31
    :::::::::::  ::  :::::::::::  ||
164 hrValCysThrCysHisAlaGlyPhePheLeuArgGluAsnGluCys... 179
    :::::::::::  ::  :::::::::::  ||
30 ATCTTCTGGCACATTTGGGTGCGATTGCAAC 1
    :::::::::::  ::  :::::::::::  ||
180 .....ValSerCysSer 183
    :::::::::::  ::  :::::::::::  ||

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-795-445A-46

alignment_scores:
  Quality: 62.00      Length: 60
  Ratio: 2.067        Gaps: 3
  Percent Similarity: 50.000  Percent Identity: 31.667

alignment_block:
  US-09-049-696-2/rev x US-08-795-445A-46 ..
  Align seg 1/1 to: US-08-795-445A-46 from: 1 to: 280

177 TTTTGGTCTCACATAGTCANCCCTTTGCTTCATGTTTCAGGAATCAAAA 128
    :::::::::::  ::  :::::::::::  ||
135 TyrTrpSerGluAsnLeuPheGlnCysPheAsnCysSerLeuCysLeuAs 151
    :::::::::::  ::  :::::::::::  ||
127 TGGCAACATTTTGAATAAATCGCTTTCCTGTAGCTTCAACAGATAC 78
    :::::::::::  ::  :::::::::::  ||
151 nGlyThrVal.....HisLeuSerCysGlnGluLysGlnAsnT 164
    :::::::::::  ::  :::::::::::  ||
77 AGAGATGCCCTGGGTCCACCATGCTTATTTGTTG...AATGAGTGTTC 31
    :::::::::::  ::  :::::::::::  ||
164 hrValCysThrCysHisAlaGlyPhePheLeuArgGluAsnGluCys... 179
    :::::::::::  ::  :::::::::::  ||
30 ATCTTCTGGCACATTTGGGTGCGATTGCAAC 1
    :::::::::::  ::  :::::::::::  ||
180 .....ValSerCysSer 183
    :::::::::::  ::  :::::::::::  ||
```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-795-445A-46

```

seq_documentation_block:
; Sequence 46, Application US/08795445A
; Patent No. 6284485
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Behavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,445A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/577,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378
; INFORMATION FOR SEQ. ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 280 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

OM of: US-09-049-696-2 to: A_Geneseq_1101.* out_format : pfs

Date: Mar 30, 2002 2:17 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL-frame+np2.model -DEV=xlp
-O/Cgn2_1/USPTO_spool/US09049696/runat_28032002_145237_2001/app_query.fasta_1.12579
-DB-A_Geneseq_1101 -OFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosome2
-TRANS-human40.cdi -LIST=45 -DICALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 @CGN1_1_293 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-2

Query length: 273

Database: A_Geneseq_1101.*

Database sequences: 522463

Database length: 74073290

Search time (sec): 819.910000

score_list:

Sequence	Strd Orig	zScore	EScore Len	Documentation
/SID2/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74822			371.00	849.05
/SID2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB73715			371.00	849.05
/SID2/gcgdata/geneseq/geneseq/AA2001.DAT: AAG75614			361.00	825.81
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAY66749			314.00	713.83
/SID2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB87560			314.00	713.83
/SID2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB65272			314.00	713.83
/SID2/gcgdata/geneseq/geneseq/AA1999.DAT: AAY41039			223.50	503.46
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAB11324			223.50	503.46
/SID2/gcgdata/geneseq/geneseq/AA1999.DAT: AAY41040			223.50	500.68
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAB11325			223.50	500.68
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAY70459			223.50	499.00
/SID2/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74823			223.50	498.99
/SID2/gcgdata/geneseq/geneseq/AA1999.DAT: AAY41036			223.50	498.99
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAB11321			223.50	498.99
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAG03643			148.00	342.41
/SID2/gcgdata/geneseq/geneseq/AA1999.DAT: AAY11789			148.00	342.31
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAG60441			67.50	142.35
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAG60440			67.50	141.39
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAG60439			67.50	139.17
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAG26241			66.50	140.78
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAG26240			66.50	140.78
/SID2/gcgdata/geneseq/geneseq/AA1997.DAT: AAW23684			66.00	139.51
/SID2/gcgdata/geneseq/geneseq/AA1990.DAT: AAR03923			66.00	132.20
/SID2/gcgdata/geneseq/geneseq/AA1990.DAT: AAR05924			66.00	132.20
/SID2/gcgdata/geneseq/geneseq/AA1999.DAT: AAW97795			66.00	132.20
/SID2/gcgdata/geneseq/geneseq/AA1999.DAT: AAG97794			65.50	138.95
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAG30416			65.50	132.20
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAB21003			65.50	125.94
/SID2/gcgdata/geneseq/geneseq/AA1999.DAT: AAY00099			64.50	132.57
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAW90780			64.50	132.35
/SID2/gcgdata/geneseq/geneseq/AA1999.DAT: AAY00098			64.50	131.92
/SID2/gcgdata/geneseq/geneseq/AA1997.DAT: AAW20051			64.50	126.12
/SID2/gcgdata/geneseq/geneseq/AA1999.DAT: AAW97097			64.50	126.12
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAY52020			64.50	126.12
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAY51649			64.50	126.12
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAG56744			64.00	135.12
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAG56743			64.00	132.45
/SID2/gcgdata/geneseq/geneseq/AA2000.DAT: AAG56742			64.00	131.85
/SID2/gcgdata/geneseq/geneseq/AA1996.DAT: AAR94397			63.50	127.42
/SID2/gcgdata/geneseq/geneseq/AA1998.DAT: AAW71488			63.50	126.97

seq_name: /SID2/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74822

seq_documentation_block:

ID AAB74822 standard; Protein; 913 AA.

XX AAB74822;

XX 13-JUN-2001 (first entry)

XX Murine ICACC-1 protein sequence.

XX ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW inflammatory bowel disease; autoimmune disease.

XX Mus sp.

XX WO9944620-A1.

XX 10-SEP-1999.

XX 03-MAR-1999; 99WO-US04703.

XX 03-MAR-1998; 98US-0076815.

XX (MAGA-) MAGAININ PHARM INC.

PI Holroyd KJ, Levitt RC, Maloy WL, Loughah J, McLane M;

PI Nicolaides NC, Zhou Y, Dong Q;

XX WPI; 1999-550979/46.

XX N-PSDB; AAF81925.

XX New nucleic acid encoding calcium activated chloride channel, used to
PT identify, e.g. specific modulators for treating atopic allergy

XX Claim 12; Fig 2; 75pp; English.

XX The present sequence represents the murine interleukin 9 (IL-9) induced
CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
CC have anti-allergic, anti-asthmatic, anti-inflammatory and
CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
CC used to alleviate asthma (or more generally atopic allergy), while those
CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
CC inflammatory bowel disease (IBD) (or other autoimmune diseases).

CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
CC (increased levels) or IBD (reduced levels), also for monitoring
CC treatment of these conditions. The ICACC proteins can be used:

CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
CC reagents, and (b) as therapeutic (A); (ii) as fragments or mutants, as
CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
CC identify modulators and binding partners. ICACC polynucleotides can be
CC used to generate transgenic animals or recombinant cells, used to screen
CC for antagonists, also as a source of therapeutic antisense agents or
CC identification of modulators).

XX Sequence 913 AA;

alignment_scores:

Quality: 371.00 Length: 91
Ratio: 4.417 Gaps: 0

Percent Similarity: 92.308 Percent Identity: 76.923

alignment_block:

us-09-049-696-2 x AAB74822

Align seg 1/1 to: AAB74822 from: 1 to: 913

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 869 AA;

alignment_scores: Quality: 361.00 Length: 75
 Ratio: 5.085 Gaps: 0
Percent Similarity: 94.667 Percent Identity: 90.667

alignment_block:

US-09-049-696-2 x AAG75614 ..

Align seg 1/1 to: AAG75614 from: 1 to: 869

49 ATAAGAGACATGGTACCCAGCAGCATCTGTATCTGTTTGAAGCTACAGG 98
||||: ||||| ||||||||| ||||||||| ||||||||| |||||||||
8 IleArgHisGluValThr**AlaSerLeuTyrLeuPheGluAlaThrGI 24
99 AAAGCGATTTTATTCAAAAGTTGCCATTTTGATTCCTGGAACATGGA 148
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
24 yLysArgPheTyrPheLysAsnValAlaIleLeuIleProGluThrTrpL 41
149 AGACAAGAGNTGACTATGTGAGACCAAACTTGAGACCTACAAAATGCT 198
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
41 yThrLysAlaAspTyrValArgProLysLeuGluThrTyrLysAsnAla 57
199 GATGTTCTGGTGTGCTGAGTCTATNTCTCCAGGNAATGATGAACCTTACAC 248
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
58 AspValLeuValAlaGluSerThrProProGlyAsnAspGluProTyrTh 74
249 TGNCGAGATGGGCACTGTGGCGAG 273
| ||||||| ||||||| |||||||
74 rGluGlnMetGlyAsnCysGlyGlu 82

seq_name: /SIDS2/jcgdata/geneseq/geneseq/AA2000.DAT:AA766749

seq_documentation_block:

ID AA766749 standard; protein; 919 AA.

XX
AC AAY66749;
DT
DT 05-APR-2000 (first entry)
XX
DE Membrane-bound protein PRO1124.
XX
KW Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW pharmaceutical; receptor immunoadhesin; gene mapping.
XX
OS Homo sapiens.
XX
PN WO9963088-A2.
XX
PD 09-DEC-1999.

XX 02-JUN-1999; 99WO-US12252.
PF 02-JUN-1998; 98US-0087607.
XX 02-JUN-1998; 98US-0087609.
PR 02-JUN-1998; 98US-0087759.
PR 03-JUN-1998; 98US-0087827.
PR 04-JUN-1998; 98US-0088021.
PR 04-JUN-1998; 98US-0088025.
PR 04-JUN-1998; 98US-0088028.
PR 04-JUN-1998; 98US-0088029.
PR 04-JUN-1998; 98US-0088030.
PR 04-JUN-1998; 98US-0088033.
PR 04-JUN-1998; 98US-0088326.
PR 05-JUN-1998; 98US-0088167.
PR 05-JUN-1998; 98US-0088202.
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.

```
DR 25-JUN-1998; 98US-0090694.
XX 25-JUN-1998; 98US-0090695.
PT 25-JUN-1998; 98US-0090696.
XX 26-JUN-1998; 98US-0090862.
PS 26-JUN-1998; 98US-0090863.
XX 01-JUL-1998; 98US-0091358.
CC 01-JUL-1998; 98US-0091360.
CC 01-JUL-1998; 98US-0091544.
CC 01-JUL-1998; 98US-0091478.
CC 02-JUL-1998; 98US-0091486.
CC 02-JUL-1998; 98US-0091519.
CC 02-JUL-1998; 98US-0091626.
CC 02-JUL-1998; 98US-0091628.
CC 02-JUL-1998; 98US-0091633.
CC 02-JUL-1998; 98US-0091646.
CC 02-JUL-1998; 98US-0091673.
CC 02-JUL-1998; 98US-0091978.
CC 07-JUL-1998; 98US-0091982.
CC 07-JUL-1998; 98US-0092182.
CC 09-JUL-1998; 98US-0092472.
CC 10-JUL-1998; 98US-0093339.
CC 20-JUL-1998; 98US-0094651.
CC 30-JUL-1998; 98US-0095282.
CC 04-AUG-1998; 98US-0095285.
CC 04-AUG-1998; 98US-0095301.
CC 04-AUG-1998; 98US-0095302.
CC 04-AUG-1998; 98US-0095318.
CC 04-AUG-1998; 98US-0095321.
CC 04-AUG-1998; 98US-0095325.
CC 04-AUG-1998; 98US-0095316.
CC 10-AUG-1998; 98US-0095929.
CC 10-AUG-1998; 98US-0096012.
CC 11-AUG-1998; 98US-0096143.
CC 11-AUG-1998; 98US-0096146.
CC 12-AUG-1998; 98US-0096329.
CC 17-AUG-1998; 98US-0096757.
CC 17-AUG-1998; 98US-0096766.
CC 17-AUG-1998; 98US-0096768.
CC 17-AUG-1998; 98US-0096773.
CC 17-AUG-1998; 98US-0096791.
CC 17-AUG-1998; 98US-0096867.
CC 17-AUG-1998; 98US-0096891.
CC 17-AUG-1998; 98US-0096894.
CC 17-AUG-1998; 98US-0096895.
CC 17-AUG-1998; 98US-0096897.
CC 18-AUG-1998; 98US-0096949.
CC 18-AUG-1998; 98US-0096950.
CC 18-AUG-1998; 98US-0096959.
CC 18-AUG-1998; 98US-0096960.
CC 18-AUG-1998; 98US-0097022.
CC 19-AUG-1998; 98US-0097141.
CC 19-AUG-1998; 98US-0097218.
CC 20-AUG-1998; 98US-0097661.
CC 24-AUG-1998; 98US-0097951.
CC 26-AUG-1998; 98US-0097952.
CC 26-AUG-1998; 98US-0097954.
CC 26-AUG-1998; 98US-0097955.
CC 26-AUG-1998; 98US-0097971.
CC 26-AUG-1998; 98US-0097974.
CC 26-AUG-1998; 98US-0097978.
CC 26-AUG-1998; 98US-0097979.
CC 26-AUG-1998; 98US-0097986.
CC 26-AUG-1998; 98US-0098014.
CC 31-AUG-1998; 98US-0098525.
CC 16-SEP-1998; 98US-0100634.
CC 12-JAN-1999; 99US-0115565.
XX
XX (GETH ) GENENTECH INC.
XX Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX WPI; 2000-072883/06.
DR

DR N-PSDB; AA265095.
XX Membrane-bound proteins and related nucleotide sequences
PT Claim 12; Fig 274; 822pp; English.
XX
XX The invention provides membrane-bound PRO polypeptides and
XX polynucleotides encoding them. The PRO sequences of the invention were
XX identified based on extracellular domain homology screening. The PRO
XX sequences have homology with proteins including LDL receptors, TIE
XX ligands and various enzymes. The membrane-bound proteins and receptor
XX molecules are useful as pharmaceutical and diagnostic agents. Receptor
XX immunoadhesins, for instance, can be used as therapeutic agents to block
XX receptor-ligand interactions. The membrane-bound proteins can also be
XX employed for screening of potential peptide or small molecule inhibitors
XX of the relevant receptor/ligand interaction. The PRO encoding sequences
XX are useful as hybridization probes, in chromosome and gene mapping and in
XX the generation of antisense RNA and DNA. PRO nucleic acid sequences
XX will also be useful for the preparation of PRO polypeptides, especially
XX by recombinant techniques.
XX
XX Sequence 919 AA:

alignment_scores:
Quality: 314.00 Length: 91
Ratio: 4.026 Gaps: 0
Percent Similarity: 85.714 Percent Identity: 64.835

alignment_block:
US-09-049-696-2 x AAY66749 ..
Align seg 1/1 to: AAY66749 from: 1 to: 919
1 GTTGCAATGACCCCAATGTGCCAGAGATGAACACTCATTCAACAAT 50
36 ILeValIleAspProSerValProGluAspGluLysIleIleGluIleI 52
51 AAAGGACATGCTGACCGACCATCTCTGTATCTGTTTGAAGCTACAGGAA 100
52 eGluAspMetValThrAlaSerThrTyrLeuPheGluAlaThrGluL 69
101 AGCGATTTTATTTCAAAATGTTGCCATTTTGTATCTCTGAAACATGGAAG 150
69 ysArgPhePhePheLysAsnValSerIleLeuIleProGluAsnTrpLys 85
151 ACAAGAGNTGACTATGTGAGACCAAACTTGAGACCTACACAAATGCTGA 200
85 GluAsnProGlnTyrLysArgProLysHisGluAsnHisLysHisAlaas 102
201 TGTTCGTGCTGAGTCTTANTCTCCAGGNAATGATGAACCCCTACACTG 250
102 pValIleValAlaProProThrLeuProGlyArgAspGluProTyrThrL 119
251 NGCAGATGGGCACTGTCGGCAG 273
119 ysGlnPheThrGluCysGlyGlu 126

seq_name: /SDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA887560
seq_documentation_block:
ID AA887560 standard; Protein; 919 AA.
XX
XX AA887560;
XX
XX 15-MAY-2001 (first entry)
XX
XX Human PRO1124.
XX
XX Human; PRO protein; mapping.
XX Homo sapiens.
XX
```


CC proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC can be used for targeted delivery of bioactive molecules, such as
CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC sequences, and their fragments, can be used as hybridisation probes, in
CC chromosomal and gene mapping, and in the generation of anti-sense RNA
CC and DNA. They may also be used to produce transgenic animals which are
CC used to develop and screen therapeutically useful reagents. The PRO
CC nucleotide and protein sequence can be used for tissue typing and in
CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
CC in the isolation of human PRO sequences. AAF44087 to AAF44269 and
CC AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC sequences given in the exemplification of the present invention.

XX
SQ Sequence 919 AA;

alignment_scores:
Quality: 314.00 Length: 91
Ratio: 4.026 Gaps: 0
Percent Similarity: 85.714 Percent Identity: 64.835
alignment_block:
US-09-049-696-2 x AAB65272 ..
Align seg 1/1 to: AAB65272 from: 1 to: 919

1 GTTGCATCGACCCCAATGTCAGAGATGAACACTCATTCAACAAAT 50
36 ILeValIleAspProSerValProGluAspGluIleIleGluGlnI 52
51 AAAGGACATGTTGACCCAGCATCTCTATCTGTTTGAAGCTACAGGAA 100
52 eGluAspMetValThrAlaSerThrTyrLeuPheGluAlaThrGluL 69
101 AGCGATTTATTCAAAATGTTGCATTTTCATTCTCTGAACATGGAAG 150
69 ysargPhePhePheLysAsnValSerIleLeuIleProGluAsnTrpLys 85
151 ACAAGGNTGACTATCTGAGACCAAACTTGAGACCTACAAAATGCTGA 200
86 GluAsnProGlnTyrLysArgProLysHisGluAsnHisLysAlaAs 102
201 TGTCTCGTTGCTGAGTCTANTCTCCAGGNAATGATGACCCCTACACTG 250
102 pValIleValAlaProProThrLeuProGlyArgAspGluProTyrThrL 119
251 NGCAGATGGCAACTGTGGCGAG 273
119 ysGlnPheThrGluCysGlyGlu 126

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AA1999
seq_documentation_block:
ID AAY41039 standard; protein; 592 AA.
XX AAY41039;
XX
XX 07-DEC-1999 (first entry)
XX Human lung tumor antigen L762P variant 1.
XX
XX Human; lung tumor; lung cancer; T cell stimulation.

XX Homo sapiens.
XX
XX W09947674-A2.
XX
XX 23-SEP-1999.
XX
XX 17-MAR-1999; 99WO-US05798.
XX
XX 18-MAR-1998; 98US-0040802.
PR

PR 18-MAR-1998; 98US-0040984.
PR 27-JUL-1998; 98US-0123912.
PR 27-JUL-1998; 98US-0123933.
XX
XX (CORI-) CORIXA CORP.
XX
XX Read SG, Wang T;
XX
XX WP: 1999-571839/48.
XX N-PSDB: AAZ24657.
XX
XX New isolated lung tumor polynucleotides, used to develop products for
XX the treatment, prevention and monitoring the progression of lung cancer
XX
XX Example 3; Page 144-145; 148pp; English.
XX
XX The invention provides isolated human lung tumor nucleic acids and
XX polypeptides. The polypeptides can be used for the treatment of lung
XX cancer. The polypeptides and polynucleotides can be used to stimulate T
XX cells or antigen presenting cells for use in the treatment of lung
XX cancer. The polypeptides and monoclonal antibodies specific for the
XX polypeptides can also be used to inhibit the development of lung cancer.
XX Agents which bind the polypeptides can be used for detecting lung cancer
XX and for monitoring the progression of lung cancer.

XX
SQ Sequence 592 AA;

alignment_scores:
Quality: 223.50 Length: 91
Ratio: 3.020 Gaps: 1
Percent Similarity: 81.319 Percent Identity: 46.154
alignment_block:
US-09-049-696-2 x AAY41039 ..
Align seg 1/1 to: AAY41039 from: 1 to: 592

1 GTTGCATCGACCCCAATGTCAGAGATGAACACTCATTCAACAAAT 50
45 ILeAlaIleAsnProGlnValProGluAsnGlnAsnLeuIleSerAsnI 61
51 AAAGGACATGTTGACCCAGCATCTCTATCTGTTTGAAGCTACAGGAA 100
61 eLysGluMetIleThrGluAlaSerPheTyrLeuPheAsnAlaThrLysA 78
101 AGCGATTTATTCAAAATGTTCCCATTTTGTCTCTGCTGAAACATGGAAG 150
78 rArgValPhePheArgAsnIleLysIleLeuIleProAlaThrTrpLys 94
151 ACAAGGNTGACTATCTGAGACCAAACTTGAGACCTACAAAATGCTGA 200
95 AlaAsnAsnAsn...SerLysIleLysGlnGluSerTyrGluLysAlaAs 110
201 TGTCTCGTTGCTGAGTCTANTCTCCAGGNAATGATGACCCCTACACTG 250
110 nValIleValThrAspTrpTyrGlyAlaHisGlyAspAspProTyrThrL 127
251 NGCAGATGGCAACTGTGGCGAG 273
127 euglnTyrArgCysGlyLys 134

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AA11324
seq_documentation_block:
ID AAB11324 standard; protein; 592 AA.
XX AAB11324;
XX
XX 21-FEB-2001 (first entry)
XX Human lung cancer-associated protein L762P variant 1.
XX
XX

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AA74823

seq_documentation_block:
ID AAB74823 standard; Protein; 943 AA.

AC AAB74823;

DT 13-JUN-2001 (first entry)

DE Human ICACC-2 protein sequence.

XX ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW inflammatory bowel disease; autoimmune disease.

XX Homo sapiens.

OS WO9944620-A1.

PN 10-SEP-1999.

PD 03-MAR-1999; 99WO-US04703.

PF 03-MAR-1998; 98US-0076815.

PR (MAGA-) MAGAININ PHARM INC..

PI Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;

PI Nicolaidis NC, Zhou Y, Dong Q;

DR WPI; 1999-550979/46.

DR N-PSDB; AAF81926.

XX New nucleic acid encoding calcium activated chloride channel, used to

PT identify, e.g. specific modulators for treating atopic allergy -

XX Claim 13; Fig 4A; 75pp; English.

XX The present sequence represents the human interleukin 9 (IL-9) induced
CC calcium activated chloride channel 2 (ICACC-2) protein. ICACC proteins
CC have anti-allergic, anti-asthmatic, anti-inflammatory and
CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
CC used to alleviate asthma (or more generally atopic fibrosis (CF) and
CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
CC (increased levels) or IBD (reduced levels), also for monitoring
CC treatment of these conditions. The ICACC proteins can be used:
CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
CC reagents, and (b) as therapeutic (A); (ii) as ligands or mutants, as
CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
CC identify modulators and binding partners. ICACC polynucleotides can be
CC used to generate transgenic animals or recombinant cells, used to screen
CC for antagonists, also as a source of therapeutic antisense agents or
CC diagnostic probes (for quantifying mRNA expression, e.g. for
CC identification of modulators).

XX Sequence 943 AA;

alignment_scores:
Quality: 223.50 Length: 91
Ratio: 3.020 Gaps: 1
Percent Similarity: 81.319 Percent Identity: 46.154

alignment_block:
US-09-049-696-2 x AAB74823

Align seg 1/1 to: AAB74823 from: 1 to: 943

1 GTTGAATGACCCCAATGTGCCAGAGATGAACACTATTCACAAAT 50

45 lleAlaIleAsnProGlnValProGlnAsnGlnAsnLeuIleSerAsnIl 61
51 AAAGACATGTGTGACCCAGGATCTCTGTATCTGTTGAAGCTACAGGAA 100
61 elysGluMetIleThrGluAlaSerPheTyrLeuPheAsnAlaThrLysA 78
101 AGCGATTTTATTCAAAAATGTGCCATTTTGTCTCTGGAACATGGAAG 150
78 rgArgValPhePheArgAsnIleLysIleLeuIleProAlaThrTriplys 94
151 ACAAGGNTGACTATGTGAGACCAAACTTGAGACCTACAAAATGCTGA 200
95 AlaAsnAsnAsn...SerLysIleLysGlnGluSerTyrGluLysAlaAs 110
201 TGTCTCTGTTGCTGAGTCTTANTCTCCAGGNAATGATGAACCTACACTG 250
110 nvalIleValThrAspTyrArgAlaHisGlyAspAspProTyrThrL 127
251 NGCAGATGGCAACTGTGCGGAG 273
127 euGlnTyrArgGlyCysGlyLys 134

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AA741036

seq_documentation_block:

ID AAY41036 standard; protein; 943 AA.

AC AAY41036;

DT 07-DEC-1999 (first entry)

DE Human lung tumor antigen L762P.

XX Human; lung tumor; lung cancer; T cell stimulation.

OS Homo sapiens.

XX WO9947674-A2.

PN 23-SEP-1999.

PD 17-MAR-1999; 99WO-US05798.

PF 18-MAR-1998; 98US-0040802.

PR 18-MAR-1998; 98US-0040984.

PR 27-JUL-1998; 98US-0123912.

PR 27-JUL-1998; 98US-0123933.

XX (CORI-) CORIXA CORP.

XX Reed SG, Wang T;

XX WPI; 1999-571839/48.

XX N-PSDB; AAZ24653.

XX New isolated lung tumor polynucleotides, used to develop products for

PT the treatment, prevention and monitoring the progression of lung cancer

XX Example 3; Page 137-139; 148pp; English.

XX The invention provides isolated human lung tumor nucleic acids and
CC polypeptides. The polypeptides can be used for the treatment of lung
CC cancer. The polypeptides and polynucleotides can be used to stimulate T
CC cells or antigen presenting cells for use in the treatment of lung
CC cancer. The polypeptides and monoclonal antibodies specific for the
CC polypeptides can also be used to inhibit the development of lung cancer.
CC Agents which bind the polypeptides can be used for detecting lung cancer
CC and for monitoring the progression of lung cancer.

XX Sequence 943 AA;

```
alignment_scores:
  Quality: 223.50      Length: 91
  Ratio: 3.020         Gaps: 1
  Percent Similarity: 81.319      Percent Identity: 46.154

alignment_block:
US-09-049-696-2 x AAY41036 ..

Align seg 1/1 to: AAY41036 from: 1 to: 943

1 GTTGCATCGACCCCAATGTGCAGAGATGAACACTCATTCAACAAT 50
  ::::::::::::::::::::|::::::::::::::::::|
45 IieAlaIleAsnProGlnValProGluAsnGlnAsnLeuIleSerAsn 61
  ::::::::::::::::::::|::::::::::::::::::|
51 AAAGACATGGTGACCCAGCATCTCTGTATCTGTTTGAAGCTACAGGA 100
  ::::::::::::::::::::|::::::::::::::::::|
61 elysGluMetIleThrGluAlaSerPheTyrLeuPheAsnAlaThrLys 78
  ::::::::::::::::::::|::::::::::::::::::|
101 AGCGATTTTATTTCAAAATGTTGCCATTTGATTCCTGAACATGAAG 150
  ::::::::::::::::::::|::::::::::::::::::|
78 rgargValPhePheArgAsnIleLysIleLeuIleProAlaThrTrpLys 94
  ::::::::::::::::::::|::::::::::::::::::|
151 ACAAGGNTGACTATCTGAGACCAAACTTGAGACCTACAAAATGCTGA 200
  ::::::::::::::::::::|::::::::::::::::::|
95 AlaAsnAsnAsn...SerLysIleLysGlnGluSerTyrGluLysAlaAs 110
  ::::::::::::::::::::|::::::::::::::::::|
201 TGTTCGTGGTGTGCTACTTANTCTCCAGGNAATGATGAACCTACACTG 250
  ::::::::::::::::::::|::::::::::::::::::|
110 nValIleValThrAspTyrGlyAlaHisGlyAspProTyrThrL 127
  ::::::::::::::::::::|::::::::::::::::::|
251 NGCAGATGGCAACTGTGGCGAG 273
  ||| ::::::::::::::::::::|
127 euGlnTyrArgGlyCysGlyLys 134
  ::::::::::::::::::::|

seq_name: /SID52/cgcdata/geneseq/geneseq/AA2000.DAT: AAB11321

seq_documentation_block:
ID AAB11321 standard; Protein: 943 AA.
XX
AC AAB11321;
XX
DT 21-FEB-2001 (first entry)
XX
DE Human lung cancer-associated protein L762S.
XX
KW Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
XX vaccine; detection.
XX
OS Homo sapiens.
XX
PN WO200061612-A2.
XX
PD 19-OCT-2000.
XX
PF 03-APR-2000; 2000WO-US08896.
XX
PR 02-APR-1999; 99US-0285479.
XX
PR 17-DEC-1999; 99US-0466396.
XX
PR 30-DEC-1999; 99US-0476496.
XX
PR 10-JAN-2000; 2000US-0480884.
XX
PR 22-FEB-2000; 2000US-0510376.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Fan L;
XX
DR WPI; 2000-628399/60.
XX
DR N-PSDB; AAC65892.
XX
PT Isolated polypeptide comprising an immunogenic portion of a lung tumor
PT protein is used for detecting and monitoring progression of lung cancer
PT in a patient -
```

```
Claim 3; Page 170-172; 261pp; English.
XX
XX This invention describes a novel isolated polypeptide (I) which
XX comprising an immunogenic portion of a lung tumor protein or variant (P2)
XX which have cytostatic activity. The polypeptides and polynucleotides are
XX used in compositions and vaccines to inhibit the development of cancer,
XX especially lung cancer, in a patient. Methods described in the invention
XX can be used to monitor the progression of a cancer by carrying out the
XX detection at subsequent time points and comparing the results from the
XX different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
XX cells expressing P2, polynucleotides encoding P2 or antigen presenting
XX development of cancer.
XX
XX Sequence 943 AA;
XX
XX alignment_scores:
XX Quality: 223.50      Length: 91
XX Ratio: 3.020         Gaps: 1
XX Percent Similarity: 81.319      Percent Identity: 46.154
XX
XX alignment_block:
XX US-09-049-696-2 x AAB11321 ..
XX
XX Align seg 1/1 to: AAB11321 from: 1 to: 943
XX
XX 1 GTTGCATCGACCCCAATGTGCAGAGATGAACACTCATTCAACAAT 50
XX ::::::::::::::::::::|::::::::::::::::::|
XX 45 IieAlaIleAsnProGlnValProGluAsnGlnAsnLeuIleSerAsn 61
XX ::::::::::::::::::::|::::::::::::::::::|
XX 51 AAAGACATGGTGACCCAGCATCTCTGTATCTGTTTGAAGCTACAGGA 100
XX ::::::::::::::::::::|::::::::::::::::::|
XX 61 elysGluMetIleThrGluAlaSerPheTyrLeuPheAsnAlaThrLys 78
XX ::::::::::::::::::::|::::::::::::::::::|
XX 101 AGCGATTTTATTTCAAAATGTTGCCATTTGATTCCTGAACATGAAG 150
XX ::::::::::::::::::::|::::::::::::::::::|
XX 78 rgargValPhePheArgAsnIleLysIleLeuIleProAlaThrTrpLys 94
XX ::::::::::::::::::::|::::::::::::::::::|
XX 151 ACAAGGNTGACTATCTGAGACCAAACTTGAGACCTACAAAATGCTGA 200
XX ::::::::::::::::::::|::::::::::::::::::|
XX 95 AlaAsnAsnAsn...SerLysIleLysGlnGluSerTyrGluLysAlaAs 110
XX ::::::::::::::::::::|::::::::::::::::::|
XX 201 TGTTCGTGGTGTGCTACTTANTCTCCAGGNAATGATGAACCTACACTG 250
XX ::::::::::::::::::::|::::::::::::::::::|
XX 110 nValIleValThrAspTyrGlyAlaHisGlyAspProTyrThrL 127
XX ::::::::::::::::::::|::::::::::::::::::|
XX 251 NGCAGATGGCAACTGTGGCGAG 273
XX ||| ::::::::::::::::::::|
XX 127 euGlnTyrArgGlyCysGlyLys 134
XX ::::::::::::::::::::|

seq_name: /SID52/cgcdata/geneseq/geneseq/AA2000.DAT: AAG03643

seq_documentation_block:
ID AAG03643 standard; Protein: 91 AA.
XX
AC AAG03643;
XX
DT 06-OCT-2000 (first entry)
XX
XX Human secreted protein, SEQ ID NO: 7724.
XX
DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PD 21-FEB-2000; 2000EP-0200610.
XX
PF
```

Tue Apr 2 09:39:55 2002

us-09-049-696-2.rag

XX 26-FEB-1999; 99US-0122487.
XX (GEST) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX N-PSDB; AAC03649.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 13; SEQ ID 7724; 71pp + CD-ROM; English.
XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX Sequence 91 AA;
SQ

alignment_scores: Quality: 148.00 Length: 46
 Ratio: 3.442 Gaps: 0
Percent Similarity: 93.478 Percent Identity: 56.522
alignment_block:
US-09-049-696-2 x AAG03643 ..
Align seg 1/1 to: AAG03643 from: 1 to: 91
1 GTTGAATCGACCCCAATGTGCCAGATGAACACTCATTCACAAT 50
 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
45 IleAlaIleAsnProGlnValProGlnAsnGlnAsnLeuIleSerAsnI 61
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 AAAGGACATGGTGACCCAGGCATCTCTGTATCTGTTTGAAGTACAGGAA 100
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 eLysGluMetIleThrGluAlaSerPheTyrLeuPheAsnAlaThrLysA 78
101 AGCGATTTTATTTCAAAAATGTTGCCATTTTGCATTCCT 138
 :|||| :|||||:|||||:|||||:|||||:|||||:|||||
78 rgArgValPhePheArgAsnIleLysIleLeuPro 90

OM of: US-09-049-696-1 to: SPTREMBL_17:* out_format : pfs

Date: Mar 30, 2002 2:45 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=framet-n2p.model -DEV=rlp
-Q/cgn2_1/USPTO-US09049696/runat_28032002_145238_2053/app_query.fasta_1.12579
-DB=SPTREMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-FCAPOP=4.500 -FCAPEXT=0.050 -XGAPOP=10.000 -XCAPEXT=0.500
-FCAPOP=6.000 -FCAPEXT=7.000 -YGAPOP=10.000 -YCAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsub2
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USRP=US09049696_@CGN1_1_705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLXPY
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-049-696-1
Query length: 223
Database: SPTREMBL_17:*
Database sequences: 473505
Database length: 14627329
Search time (sec): 805.760000
```

score_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
sp_human:O95151	+	328.00	718.72	1.0e-32	914	! O95151 homo sapiens (human). ca
sp_human:Q9UPC6	+	328.00	718.72	1.0e-32	914	! Q9UPC6 homo sapiens (human). ca
sp_mammal:Q9TUB5	+	278.00	605.79	2.0e-26	917	! Q9TUB5 sus scrofa (pig). epith
sp_rodent:Q68826	+	257.00	558.41	8.6e-24	913	! Q68826 mus musculus (mouse). gc
sp_rodent:Q9D726	+	257.00	558.41	8.6e-24	913	! Q9D726 mus musculus (mouse). ch
sp_human:Q9UNF7	+	196.50	401.76	3.5e-16	917	! Q9UNF7 homo sapiens (human). ca
sp_rodent:Q88860	+	187.50	401.57	4.8e-15	902	! Q88860 mus musculus (mouse). ch
sp_rodent:Q9R070	+	187.50	401.57	4.8e-15	902	! Q9R070 mus musculus (mouse). ca
sp_rodent:Q9QX15	+	187.50	401.57	4.8e-15	902	! Q9QX15 mus musculus (mouse). ca
sp_rodent:Q9E0R4	+	187.50	401.57	4.8e-15	902	! Q9E0R4 mus musculus (mouse). en
sp_human:Q9F6N3	+	184.00	403.00	1.4e-14	262	! Q9F6N3 homo sapiens (human). cl
sp_mammal:Q18744	+	181.50	395.34	2.8e-14	342	! Q18744 bos taurus (bovine). lu
sp_mammal:Q18742	+	181.50	388.98	2.7e-14	794	! Q18742 bos taurus (bovine). lu
sp_mammal:Q18743	+	181.50	388.74	2.7e-14	820	! Q18743 bos taurus (bovine). lu
sp_mammal:Q18741	+	181.50	387.99	2.7e-14	905	! Q18741 bos taurus (bovine). lu
sp_human:Q9UQC9	+	148.50	313.17	3.8e-10	943	! Q9UQC9 homo sapiens (human). ch
sp_human:Q9F6N2	+	148.50	313.17	3.8e-10	943	! Q9F6N2 homo sapiens (human). ca
sp_bacteria:Q9CPA1	+	79.00	164.73	0.2184	306	! Q9CPA1 pasteurella multocida. h
sp_organelle:Q9XKX6	-	73.00	155.44	1.26	174	! Q9XKX6 chryseymys picta. nadh de
sp_organelle:Q9MR95	-	70.00	153.82	3.08	88	! Q9MR95 chelonina mydas caranigra
sp_organelle:Q9XPH5	-	70.00	148.67	3.01	174	! Q9XPH5 plasmodium falciparum
sp_invertebrate:Q97293	+	70.00	139.11	2.89	617	! Q97293 plasmodium falciparum
sp_invertebrate:Q9V637	+	67.50	139.71	6.13	270	! Q9V637 drosophila melanogaste
sp_virus:P88403	+	66.00	130.66	9.48	258	! P88403 human immunodeficiency v
sp_plant:Q9CAH2	+	66.00	126.49	9.08	993	! Q9CAH2 arabidopsis thaliana (mc
sp_invertebrate:Q77310	+	65.50	122.68	10.38	1417	! Q77310 plasmodium falciparum
sp_bacteria:Q9RFG7	+	64.50	131.92	14.55	309	! Q9RFG7 vibrio harveyi. group i
sp_mammal:Q9N0X3	+	63.50	134.91	19.87	154	! Q9N0X3 ovis aries (sheep). kuni
sp_invertebrate:Q9XXN1	+	63.50	124.66	19.03	599	! Q9XXN1 caenorhabditis elegans
sp_human:Q9NYG5	-	62.50	137.23	27.06	84	! Q9NYG5 homo sapiens (human). hep
sp_rodent:Q9CPX9	-	62.50	137.23	27.06	84	! Q9CPX9 mus musculus (mouse). lli
sp_human:Q9P0R2	-	62.50	133.82	26.67	132	! Q9P0R2 homo sapiens (human). hs
sp_bacteria:Q07842	-	62.50	126.59	25.87	344	! Q07842 rhodobacter sphaeroides
sp_invertebrate:Q9VKY2	+	62.50	116.86	24.84	1248	! Q9VKY2 drosophila melanogaste
sp_human:Q9H616	-	62.00	129.33	30.39	206	! Q9H616 homo sapiens (human). cd
sp_virus:P88390	+	62.00	128.27	30.26	237	! P88390 human immunodeficiency v
sp_rodent:Q9DC08	+	62.00	121.19	29.37	606	! Q9DC08 mus musculus (mouse). wd
sp_rodent:Q9DBM8	+	62.00	121.19	29.37	606	! Q9DBM8 mus musculus (mouse). wc
sp_organelle:Q47955	+	62.00	120.33	29.27	679	! Q47955 trypanosoma brucei. hypo

```
sp_invertebrate:Q9VAY8 + 61.50 124.40 34.57 341 ! Q9VAY8 drosophila melanoga
sp_invertebrate:Q9VZ33 + 61.50 122.54 34.30 436 ! Q9VZ33 drosophila melanoga
sp_plant:Q9ZTY7 - 61.50 122.20 34.26 456 ! Q9ZTY7 solanum berthaultii.
sp_bacteria:Q00935 + 61.50 118.23 33.69 772 ! Q00935 bacteroides thetaiota
sp_human:Q9UG05 + 61.00 126.11 40.44 234 ! Q9UG05 homo sapiens (human).
```

seq_name: sp_human:Q95151

seq_documentation_block:

```
ID O95151 PRELIMINARY; PRT; 914 AA.
AC O95151;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBL_TaxID=9606;
FN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Elble R.C., Ji H.L., Schreur K.D., Fuller C.M.,
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
RT of human CLCAL, the first human member of the family of Ca2+-activated
RT Cl-channel proteins."
RL Genomics 54:200-214(1998).
DR EMBL; AF039400; AAC95428.1; -
DR InterPro; IPR000131; ATPase_gamma.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS02334; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR SEQUENCE 914 AA; 100174 MW; 7066440C46557FA3 CRC64;
```

```
alignment_scores:
  Quality: 328.00      Length: 66
  Ratio: 4.970        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-1 x O95151

Align seg 1/1 to: O95151 from: 1 to: 914
```

```
25 ATGGGGCCATTTAAAGAGTTCTGTGTCATCTTGCATCTTTCACCTTCTTAGA 74
|||||
1 MetGlyProPhelySerSerValPheLeuLeuLeuLeuLeuLeuLeuLeuLeu 17
75 AGGGGCCCTGACGTAATTCATCTCAGCTGACCAACAAATGGCTATGAAG 124
|||||
17 uGlyAlaLeuSerAsnSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 34
125 GCATTGTCTGTCATTCGACCCCAATGTCGAGAGATGAGAACTCATTT 174
|||||
34 LytleValValAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 50
175 CAACAAATAAGGACATGTCGAGCCAGGATCTCTGTCATCTGTTTCAA 222
|||||
51 GlnGlnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 66
```

seq_name: sp_human:Q9UPC6

seq_documentation_block:

```
ID Q9UPC6 PRELIMINARY; PRT; 914 AA.
AC Q9UPC6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
```

```

DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Elble R.C., Ji H.L., Schreier K.D., Fuller C.M.,
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
RT of human CLCAL, the first human member of the family of Ca2+-activated
RT Cl-channel proteins."
RL Genomics 54:200-214(1998).
DR EMBL; AF039401; AAC95429.1; -.
DR InterPro; IPR00131; ATPase_gamma.
DR PROSITE; PS00153; ATPase_GAMMA; UNKNOWN_1.
DR PROSITE; PS00327; VWFA; 1.
DR SMART; SM00327; VWFA; 1.
SQ SEQUENCE 914 AA; 100160 MW; 7761362B66752AA3 CRC64;

alignment_scores:
Quality: 328.00 Length: 66
Ratio: 4.970 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-1 x Q9UPC6
Align seg 1/1 to: Q9UPC6 from: 1 to: 914

25 ATGGGGCCATTAAAGAGTTCTGTTCATCTTGTATCTTCCACCTCTCTAGA 74
|||||
1 MetGlyProPheLysSerValPheLeuLeuLeuHisLeuLeuGlu 17

75 AGGGGCCCTGAGTAATTCATCTACCTGACCTGAACACAAATGGCTATGAAG 124
|||||
17 uGlyAlaLeuSerAsnSerLeuLeuGlnLeuAsnAsnGlyTyrGluG 34

125 GCATTGTCTGTCATCAATGACCCCAATGTGCCAGAGATGAACACTCATTT 174
|||||
34 lylleValValAlaIleAspProAsnValProGluAspGluThrLeuLe 50

175 CAACAATAAGGACATGGTGACCCAGCATCTCTGTATCTGTTTGA 222
|||||
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGlu 66

seq_name: sp_human:Q9UPC6

seq_documentation_block:
ID Q9UPC6 PRELIMINARY; PRT; 914 AA.
AC Q9UPC6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL PROTEIN 1.
GN CACCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=SMALL INTESTINE, AND COLON;
RA MEDLINE=99364503; PubMed=10437792;
RA Aguel M., Vermet T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
RT chloride channel (CaCC) family predominantly expressed in the
RT digestive tract and trachea."
RL FEBS Lett. 455:295-301(1999).

```

```

DR EMBL; AF127036; AAD25487.1; -.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS00153; ATPase_GAMMA; UNKNOWN_1.
DR PROSITE; PS00234; VWFA; 1.
SQ SEQUENCE 914 AA; 100202 MW; B58B451D97887D6B CRC64;

alignment_scores:
Quality: 328.00 Length: 66
Ratio: 4.970 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-1 x Q9UNF6
Align seg 1/1 to: Q9UNF6 from: 1 to: 914

25 ATGGGGCCATTAAAGAGTTCTGTTCATCTTGTATCTTCCACCTCTCTAGA 74
|||||
1 MetGlyProPheLysSerValPheLeuLeuLeuHisLeuLeuGlu 17

75 AGGGGCCCTGAGTAATTCATCTACCTGACCTGAACACAAATGGCTATGAAG 124
|||||
17 uGlyAlaLeuSerAsnSerLeuLeuGlnLeuAsnAsnGlyTyrGluG 34

125 GCATTGTCTGTCATCAATGACCCCAATGTGCCAGAGATGAACACTCATTT 174
|||||
34 lylleValValAlaIleAspProAsnValProGluAspGluThrLeuLe 50

175 CAACAATAAGGACATGGTGACCCAGCATCTCTGTATCTGTTTGA 222
|||||
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGlu 66

seq_name: sp_mammal:Q9UNF6

seq_documentation_block:
ID Q9UNF6 PRELIMINARY; PRT; 917 AA.
AC Q9UNF6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE EPIHELIAL CHLORIDE CHANNEL PROTEIN.
GN AECC.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=ILEAL MUCOSA;
RA Gaspar K.J., Gabriel S.E., Racette K.J., Forsyth G.W.;
RT "A cDNA involved in porcine exocrine chloride conductance."
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF095584; AAF00077.1; -.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPase_GAMMA; UNKNOWN_1.
DR PROSITE; PS00234; VWFA; 1.
DR SMART; SM00327; VWFA; 1.
SQ SEQUENCE 917 AA; 100735 MW; 649F41944F7EAD19 CRC64;

alignment_scores:
Quality: 278.00 Length: 66
Ratio: 4.484 Gaps: 0
Percent Similarity: 93.939 Percent Identity: 83.333

alignment_block:
US-09-049-696-1 x Q9TUB5
Align seg 1/1 to: Q9TUB5 from: 1 to: 917

```



```

25 ATGGGGCCATTAAAGATTCTGTGTTTCATCTTCATTTCTCACCTTCTAG 74
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetGlySerPheArgSerLeuPheLeuValLeuHisLeuLeuG1 17
75 AGGGGCCCTGAGTAATTCATCTACCTGACGTAACAACTGCGTATGAAG 124
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 uGlyAlaGlnSerAsnSerLeuIleGlnLeuAsnGlyAsnGlyTyrgluG 34
125 GCATTGCTGTTGCAATCGACCCCAATGTGCCAGAGATGAAACACTCAT 174
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 lyileValIleAlaIleAspProAsnValProGluAspGluArgLeuLeu 50
175 CAACAATAAAGACATGCTGACCCAGGATCTCTGTATCTGTTGAA 222
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 GlnHisIleLysAspMetValThrLysAlaSerProTyrgluPheGlu 66

```

seq_name: sp_rodent:088826

```

seq_documentation_block:
ID O88826 PRELIMINARY; PRT; 913 AA.
AC O88826;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE GOB-5 PROTEIN.
DE CLCA3 OR GOB-5.
GN CLCA3 OR GOB-5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RA Komiya T., Tanigawa Y., Hirohashi S.;
RT "Cloning of a gene, gob-5, which is expressed in intestinal goblet
cells in mice."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017156; BAA33743.1; -.
DR MGD; MGI:1346342; Clca3.
DR InterPro; IPR00131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
SQ SEQUENCE 913 AA; 100070 MW; A7FA2F9E1089806D CRC64;

```

```

alignment_scores:
Quality: 257.00 Length: 66
Ratio: 4.213 Gaps: 0
Percent Similarity: 92.424 Percent Identity: 80.303
alignment_block:
US-09-049-696-1 x O88826
Align seg 1/1 to: O88826 from: 1 to: 913

```

```

25 ATGGGGCCATTAAAGATTCTGTGTTTCATCTTCATTTCTCACCTTCTAG 74
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetGlySerLeuLysSerProValPheLeuLeuLeuHisLeuLeuG1 17
75 AGGGGCCCTGAGTAATTCATCTACCTGACGTAACAACTGCGTATGAAG 124
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 uGlyValLeuSerGluSerLeuIleGlnLeuAsnAsnGlyTyrgluG 34
125 GCATTGCTGTTGCAATCGACCCCAATGTGCCAGAGATGAAACACTCAT 174
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 lyileValIleAlaIleAspHisAspValProGluAspGluAlaLeuLeu 50
175 CAACAATAAAGACATGCTGACCCAGGATCTCTGTATCTGTTGAA 222
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 GlnHisIleLysAspMetValThrGlnAlaSerProTyrgluPheGlu 66

```

seq_name: sp_rodent:Q9D7z6

```

seq_documentation_block:
ID Q9D7z6 PRELIMINARY; PRT; 913 AA.
AC Q9D7z6;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CHLORIDE CHANNEL CALCIUM ACTIVATED 3.
GN CLCA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=STOMACH;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Suzuki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Wyszewski H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszewski H., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK008659; BAB25815.1; -.
DR MGD; MGI:1346342; Clca3.
DR InterPro; IPR00131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
SQ SEQUENCE 913 AA; 100111 MW; 491E8584B06D9A89 CRC64;

```

```

alignment_scores:
Quality: 257.00 Length: 66
Ratio: 4.213 Gaps: 0
Percent Similarity: 92.424 Percent Identity: 80.303
alignment_block:
US-09-049-696-1 x Q9D7z6
Align seg 1/1 to: Q9D7z6 from: 1 to: 913

```

```

25 ATGGGGCCATTAAAGATTCTGTGTTTCATCTTCATTTCTCACCTTCTAG 74
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetGlySerLeuLysSerProValPheLeuLeuLeuHisLeuLeuG1 17
75 AGGGGCCCTGAGTAATTCATCTACCTGACGTAACAACTGCGTATGAAG 124
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 uGlyValLeuSerGluSerLeuIleGlnLeuAsnAsnGlyTyrgluG 34
125 GCATTGCTGTTGCAATCGACCCCAATGTGCCAGAGATGAAACACTCAT 174
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 lyileValIleAlaIleAspHisAspValProGluAspGluAlaLeuLeu 50
175 CAACAATAAAGACATGCTGACCCAGGATCTCTGTATCTGTTGAA 222
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


Percent Similarity: 86.364 Percent Identity: 57.576

alignment_block:

US-09-049-696-1 x Q9R070 ..

Align seg 1/1 to: Q9R070 from: 1 to: 902

```

25 ATGGGGCCCATTTAAGAGTCTGTGTTTCATCTGATTCACCTTCTCTAGA 74
   ||| ||| : : : : : ||| ||| ||| ||| ||| ||| ||| |||
1 MetValProGlyLeuGlnValLeuLeuPheLeuThrLeuHisLeuLeuG1 17
   : : : : : : : : : : : : : : : : : : : : : : : :
75 AGGGGCCCTGAGTAATTCACCTCAGCTGACGACAAACATGGCTATGAAG 124
   : : : : : : : : : : : : : : : : : : : : : : : :
17 nAsnThr...GlusSerMetValHisLeuAsnSerAsnGlyTyrGluG 33
   : : : : : : : : : : : : : : : : : : : : : : : :
125 GCATTGCTGCTGCAATCGACCCCAATGTGCCAGAGATGAACACATTCATT 174
   ||| ||| : : : : : ||| ||| ||| ||| ||| ||| ||| |||
33 lyValValIleAlaIleAsnProSerValProGluAspGluArgLeuIle 49
   : : : : : : : : : : : : : : : : : : : : : : : :
175 CAACAATAAAGACATGTTGACCCAGGACATCTCTGATCTGTTTGAAG 222
   : : : : : : : : : : : : : : : : : : : : : : : :
50 ProSerIleTysGluMetValThrGlnAlaSerThrTyrLeuPheGlu 65

```

seq_name: sp_rodent:Q9QX15

seq_documentation_block:

```

ID Q9QX15 PRELIMINARY; PRT; 902 AA.
AC Q9QX15;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIUM-SENSITIVE CHLORIDE CONDUCTANCE PROTEIN-1.
GN CLCA1 OR MCLCAL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=WHOLE LUNG;
RX MEDLINE=99041980; PubMed=9822685;
RA Gandhi R., Elble R.C., Gruber A.D., Schreur K.D., Ji H.-L.,
RA Fuller C.M., Pauli B.U.;
RT "Molecular and functional characterization of a calcium-sensitive
chloride channel from mouse lung.";
RL J. Biol. Chem. 273:32096-32101(1998).
DR EMBL; AF047838; AAC79982.1; -.
DR MGD; MGI:1316732; Clcal.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 902 AA; 100107 MW; D9A915E698EA2D85 CRC64;

```

alignment_scores:

Quality: 187.50 Length: 66
Ratio: 3.289 Gaps: 1
Percent Similarity: 86.364 Percent Identity: 57.576

alignment_block:

US-09-049-696-1 x Q9QX15 ..

Align seg 1/1 to: Q9QX15 from: 1 to: 902

```

25 ATGGGGCCCATTTAAGAGTCTGTGTTTCATCTGATTCACCTTCTCTAGA 74
   ||| ||| : : : : : ||| ||| ||| ||| ||| ||| ||| |||
1 MetValProGlyLeuGlnValLeuLeuPheLeuThrLeuHisLeuLeuG1 17
   : : : : : : : : : : : : : : : : : : : : : : : :
75 AGGGGCCCTGAGTAATTCACCTCAGCTGACAAACATGGCTATGAAG 124
   : : : : : : : : : : : : : : : : : : : : : : : :
17 nAsnThr...GlusSerMetValHisLeuAsnSerAsnGlyTyrGluG 33
   : : : : : : : : : : : : : : : : : : : : : : : :
125 GCATTGCTGCTGCAATCGACCCCAATGTGCCAGAGATGAACACATTCATT 174

```

```

||| ||| : : : : : ||| ||| ||| ||| ||| ||| ||| |||
33 lyValValIleAlaIleAsnProSerValProGluAspGluArgLeuIle 49
   : : : : : : : : : : : : : : : : : : : : : : : :
175 CAACAATAAAGACATGTTGACCCAGGACATCTCTGATCTGTTTGAAG 222
   : : : : : : : : : : : : : : : : : : : : : : : :
50 ProSerIleTysGluMetValThrGlnAlaSerThrTyrLeuPheGlu 65

```

seq_name: sp_rodent:Q9EQR4

seq_documentation_block:

```

ID Q9EQR4 PRELIMINARY; PRT; 902 AA.
AC Q9EQR4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ENDOTHELIAL CHLORIDE CHANNEL.
GN CLCA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57/BL6;
RA Elble R.C., Pauli B.;
RT "A murine endothelial homolog of Lu-ECAM-1 associated with lung
melanoma metastasis.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF115852; AAG47626.1; -.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00080; FN3; 1.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 902 AA; 99855 MW; F2BA57DAA5D4AAAD CRC64;

```

alignment_scores:

Quality: 187.50 Length: 66
Ratio: 3.289 Gaps: 1
Percent Similarity: 86.364 Percent Identity: 57.576

alignment_block:

US-09-049-696-1 x Q9EQR4 ..

Align seg 1/1 to: Q9EQR4 from: 1 to: 902

```

25 ATGGGGCCCATTTAAGAGTCTGTGTTTCATCTGATTCACCTTCTCTAGA 74
   ||| ||| : : : : : ||| ||| ||| ||| ||| ||| ||| |||
1 MetValProGlyLeuGlnValLeuLeuPheLeuThrLeuHisLeuLeuG1 17
   : : : : : : : : : : : : : : : : : : : : : : : :
75 AGGGGCCCTGAGTAATTCACCTCAGCTGACAAACATGGCTATGAAG 124
   : : : : : : : : : : : : : : : : : : : : : : : :
17 nAsnThr...GlusSerMetValHisLeuAsnSerAsnGlyTyrGluG 33
   : : : : : : : : : : : : : : : : : : : : : : : :
125 GCATTGCTGCTGCAATCGACCCCAATGTGCCAGAGATGAACACATTCATT 174
   ||| ||| : : : : : ||| ||| ||| ||| ||| ||| ||| |||
33 lyValValIleAlaIleAsnProSerValProGluAspGluArgLeuIle 49
   : : : : : : : : : : : : : : : : : : : : : : : :
175 CAACAATAAAGACATGTTGACCCAGGACATCTCTGATCTGTTTGAAG 222
   : : : : : : : : : : : : : : : : : : : : : : : :
50 ProSerIleTysGluMetValThrGlnAlaSerThrTyrLeuPheGlu 65

```

seq_name: sp_human:Q9Y6N3

seq_documentation_block:

```

ID Q9Y6N3 PRELIMINARY; PRT; 262 AA.
AC Q9Y6N3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE ENDOTHELIAL CHLORIDE CHANNEL.
GN CLCA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=WHOLE LUNG;
RX MEDLINE=99041980; PubMed=9822685;
RA Gandhi R., Elble R.C., Gruber A.D., Schreur K.D., Ji H.-L.,
RA Fuller C.M., Pauli B.U.;
RT "Molecular and functional characterization of a calcium-sensitive
chloride channel from mouse lung.";
RL J. Biol. Chem. 273:32096-32101(1998).
DR EMBL; AF047838; AAC79982.1; -.
DR MGD; MGI:1316732; Clcal.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 262 AA; 100107 MW; D9A915E698EA2D85 CRC64;

```

Align seq 1/1 to: O18744 from: 1 to: 342

```
46 GTGTTCACTCTTATTCTTCACTCTTAGAGGGCCCTGAGTAATTCAC 95
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
8  ILeuPheLeuThrLeuHisLeuLeuProGly...MetLysSerSerMe 23
   : : : : : : : : : : : : : : : : : : : : : : : : : :
96 CATTCACTCAACAACAATGGCATTCAAGGCATTCGTTCGAATCGACC 145
   : : : : : : : : : : : : : : : : : : : : : : : : : :
23 tValAsnLeuIleAsnAsnGlyThrAspGlyIleValIleAlaIleAsn 40
   : : : : : : : : : : : : : : : : : : : : : : : : : :
146 CCAATGTGCAGAGATGAACAACCTCATTCACAAATAAAGGACATGGTG 195
   : : : : : : : : : : : : : : : : : : : : : : : : : :
40 roserValProGluAspGluLeuLysIleGluAsnIleLysGluMetVal 56
   : : : : : : : : : : : : : : : : : : : : : : : : : :
```

196 ACCCAGGCATCTCTGTATCTGTTT 219
|||:::|||||
57 ThrGluAlaSerThrTyrLeuphe 64

seq_name: sp_mammal:O18742

see documentation block:

ID	018742	PRELIMINARY DOCUMENTATION BLOCK.
AC	018742;	
DT	01-JAN-1998	(TREMBLE)

DT	01-JAN-1998	(TREMBLER)
DT	01-JUN-2001	(TREMBLER)

DE LU-ECAM-I.
OS Bos taurus (Bovine).
OC Eubacterium

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Carnivora; Canidae; Bovidae; Bovinae; Bos taurus

```

OC
OX
OX
RN
-----
DOVIDUE, BOVINUE, BOS
NCBI_TaxID=9913;
[]

```

RC | SEQUENCE FROM N.A.
RC | TISSUE=LUNG;

RA Elble R.C., Widom J.,
RA Goodwin A., Pauli B.U.

Submitted (APR-1997) to
EMBL; AF001262; AAB865

DR	InterPro; IPR002035; V
DR	FROSITE; PS50234; VWF

DR	SMART; SM00327; VWA; 1
SQ	SEQUENCE 794 AA; 88

— — — — —

alignment_scores:	
Quality:	181.5
Part:0	2400

Percent Similarity:	89.655
Ratio:	3.7490

```
alignment_block:
US-09-049-696-1 v 018742
```

Align seg 1/1 to: 018742

46 GTGTTCACTTGATTCTT

8 IleLeuPheLeuThrLeu

96 CATTCAGCTGAACAACAA

```

|:::| | | |
23 tValAsnLeuIleAsnAs

```

146 CCAATGTGCCAGAAGATG

```

40 roSerValProGluAspG
|:::|

```

196 ACCCAGGCATCTCTGTAT

57 ThrGluAlaSerThrTyr

seq_name: sp_mammal:018743

2

10

```
seq_documentation_block:
ID   O18743      PRELIMINARY;          PRT;   820 AA.
AC   O18743;
DT   01-JAN-1998 (TReMBLrel. 05, Created)
DT   01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE   LU-ECAM-1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG;
RA   Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA   Goodwin A., Pauli B.U.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR   EMBL; AF001263; AAB86531.1; -.
DR   InterPro; IPR002035; VMFA.
DR   PROSITE; PS50234; VMFA; 1.
DR   SMART; SM00327; VWA; 1.
SQ   SEQUENCE 820 AA; 91464 MW; 4E01330AEF5A1494 CRC64;
```

```
alignment_scores:
  Quality: 181.50      Length: 58
  Ratio: 3.490        Gaps: 1
  Percent Similarity: 89.655      Percent Identity: 60.345

alignment_block:
US-09-049-696-1 x O18743  ..

Align seg 1/1 to: O18743 from: 1 to: 820

46  GTGTTTCATCTTGATTCTTCACCTTCTAGAAGGCCCTCGAGTAATTCAC 95
8   ILeuPheLeuThrLeuHisLeuLeuProGly...MetLysSerSerme 23
96  CATTCAGCTGAACAACAATGGCTATGAAGGCATTTGCTGTGCAATCGACC 145
23  tValAsnLeuIleAsnAsnGlyTyrAspGlyIleValIleAlaIleAsn 40
146 CCAATGTGCCAGAGATGAACACTCATTCACACAATAAAGGACATGGTG 195
40  roSerValProGluAspGluLysLeuIleGluAsnIleLysGluMetVal 56
196 ACCGAGCATCTCTGTATCTGTTT 219
57  ThrGluAlaSerThrTyrLeuPhe 64
```

This Page Blank (uspto)


```
Ratio: 3.409 Caps: 1
Percent Similarity: 84.615 Percent Identity: 58.462

alignment_block:
US-09-049-696-1 x ECLC_BOVIN
Align seg 1/1 to: ECLC_BOVIN from: 1 to: 903

25 ATGGGGCCATTAAAGATCTGCTGTCATCTTACATCTTACCTTCTAG 74
   ||| ||| ::::::::::::::| | | | | | | | | | | |
1 MetValProArgLeuThrValIleLeuPheLeuThrLeuHisLeuLeuPr 17

75 AGGGCCCTTGACTAATTCACCTATTCACCTGAACAACAATGGCTATGA 124
   ||| ||| ::::::::::::::| | | | | | | | | | | |
17 cdy...MetLysSerMetValAsnLeuIleAsnGlyTyrAsp 33

125 GCATTGCTGTCGATGACCCCAATGTCAGAGAGATGAACACTCAT 174
   ||| ||| ::::::::::::::| | | | | | | | | | | |
33 lylleValIleAlaIleAsnProSerValProGluAspGluLysLeu 49

175 CACAAATAAGGACATGTCGACCCAGGATCTCTGTATCTGTTT 219
   ||| ||| ::::::::::::::| | | | | | | | | | | |
50 GlnAsnIleLysGluMetValThrGluAlaSerThrTyrLeuPhe 64

seq_name: SwissProt_39:UL71_HSV60

seq_documentation_block:
ID UL71_HSV60 STANDARD; PRT; 213 AA.
AC P52473;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN U44.
GN U44.
OS Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Roseolovirus.
OX NCBI_TaxID=10370;
RN [1]
RP MEDLINE=95286321; PubMed=7747482;
RA Compels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,
RA Martin M.E., Efsthaliou S., Craxton M., Macaulay H.A.;
RT "The DNA sequence of human herpesvirus-6: structure, coding content,
RT and genome evolution."
RL Virology 209:29-51(1995).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS TOGETHER EBV BSF1,
CC HSV-6 U44, HVS-1 55 AND HCMV UL71.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X83413; CAAS8378.1;
DR EMBL; X92436; CAAG3170.1;
SQ SEQUENCE 213 AA; 24296 MW; 6450CD865A786D7 CRG64;

alignment_scores:
Quality: 66.00 Length: 39
Ratio: 2.276 Gaps: 0
Percent Similarity: 74.359 Percent Identity: 33.333

alignment_block:
US-09-049-696-1 x UL71_HSV60
Align seg 1/1 to: UL71_HSV60 from: 1 to: 213

88 AATTCACTCATTCAGCTGAACAACAATGGCTATGAAGCATTCGTTGC 137
   ||| ||| ::::::::::::::| | | | | | | | | | | |
1 MetValProArgLeuThrValIleLeuPheLeuThrLeuHisLeuLeuPr 17

175 AGGGCCCTTGACTAATTCACCTATTCACCTGAACAACAATGGCTATGA 124
   ||| ||| ::::::::::::::| | | | | | | | | | | |
17 cdy...MetLysSerMetValAsnLeuIleAsnGlyTyrAsp 33

125 GCATTGCTGTCGATGACCCCAATGTCAGAGAGATGAACACTCAT 174
   ||| ||| ::::::::::::::| | | | | | | | | | | |
33 lylleValIleAlaIleAsnProSerValProGluAspGluLysLeu 49

175 CACAAATAAGGACATGTCGACCCAGGATCTCTGTATCTGTTT 219
   ||| ||| ::::::::::::::| | | | | | | | | | | |
50 GlnAsnIleLysGluMetValThrGluAlaSerThrTyrLeuPhe 64

seq_name: SwissProt_39:STCK_EMENI

seq_documentation_block:
ID STCK_EMENI STANDARD; PRT; 1914 AA.
```

```
||||| ||| ::::::::::::::| | | | | | | | | | | |
158 AsnSerLeuIleGluIleAspGlyGluAsnSerLysArgLeuLeuValG 174
   ::| | | | | | | | | | | | | | | | | | | | | |
138 AATGACCCCAATGTCAGAGAGATGAACAACAATTCATTCACAAATAAAG 187
   ::| | | | | | | | | | | | | | | | | | | | | |
174 uLeuAspProIleLeuHisGluGluThrGlyLeuTyrGlnAlaLeuProA 191

188 ACATGGTGACCCAGGCA 204
   ::| | | | | | | | | | | | | | | | | | | | | |
191 snValValThrGluAla 196

seq_name: SwissProt_39:UL71_HSV6Z

seq_documentation_block:
ID UL71_HSV6Z STANDARD; PRT; 231 AA.
AC P52546;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN U44.
GN U44 OR KA5R.
OS Human herpesvirus (type 6 / strain 229) (HHV6).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Roseolovirus.
OX NCBI_TaxID=36351;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95074921; PubMed=7983761;
RA Stamey F.R., Dominguez G., Black J.B., Dambaugh T.R., Pellett P.E.;
RT "Intragenomic linear amplification of human herpesvirus 68 oriLyt
RT suggests acquisition of oriLyt by transposition."
RL J. Virol. 69:589-596(1995).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS TOGETHER EBV BSF1,
CC HSV-6 U44, HVS-1 55 AND HCMV UL71.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF157706; AAB06342.1;
DR EMBL; AF157706; AAB06342.1;
SQ SEQUENCE 231 AA; 26151 MW; 53B3EE29C16E87F2 CRG64;

alignment_scores:
Quality: 66.00 Length: 39
Ratio: 2.276 Gaps: 0
Percent Similarity: 74.359 Percent Identity: 33.333

alignment_block:
US-09-049-696-1 x UL71_HSV6Z
Align seg 1/1 to: UL71_HSV6Z from: 1 to: 231

88 AATTCACCTCATTCAGCTGAACAACAATGGCTATGAAGCATTCGTTGC 137
   ||| ||| ::::::::::::::| | | | | | | | | | | |
158 AsnSerLeuIleGluIleAspGlyGluAsnSerLysArgLeuLeuValG 174

138 AATGACCCCAATGTCAGAGAGATGAACAACAATTCATTCACAAATAAAG 187
   ::| | | | | | | | | | | | | | | | | | | | | |
174 uLeuAspProIleLeuHisGluGluThrGlyLeuTyrGlnAlaLeuProA 191

188 ACATGGTGACCCAGGCA 204
   ::| | | | | | | | | | | | | | | | | | | | | |
191 snValValThrGluAla 196

seq_name: SwissProt_39:STCK_EMENI

seq_documentation_block:
ID STCK_EMENI STANDARD; PRT; 1914 AA.
```


FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 507 AA; 56614 MW; 0BAFA046B9E85AC5 CRC64;

alignment_scores:
Quality: 64.00 Length: 38
Ratio: 2.462 Gaps: 1
Percent Similarity: 68.421 Percent Identity: 34.211

alignment_block:
US-09-049-696-1/rev x SXA2_SCHPO ..

Align seg 1/1 to: SXA2_SCHPO from: 1 to: 507

222 TTCACACAGATACAGAGATGC.....CTGGTCACCATGCTCT 185
|||||:|||||: ||| |||:|||||: |||
322 PheAsnSerIleSerGlyCysAspLeuTyrSerLeuSerAsnPheLeu 338
184 TTATTGTGTAAGTACTTTTCATCTTCGGCACATTCGGGTGCAATTGCA 135
|||||:|||||: ||| |||: |||: |||: |||: |||
338 uTyrLeuGluAsnSerCysValIleThrTyrAspValSerLeuAspCys 355
134 ACCACAATGCCCTTC 121
:: ||| ::
355 erPheAsnGluTyr 359

seq_name: SwissProt_39:RRPP_TPMV

seq_documentation_block:
ID RRPP_TPMV STANDARD; PRT; 527 AA.
AC Q9WS39;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RNA POLYMERASE ALPHA SUBUNIT (EC 2.7.7.48) (NUCLEOCAPSID
DE PHOSPHOPROTEIN).
GN P/V/C.
OS Tupala paramyxovirus (TPMV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Paramyxovirus.
OX NCBI_TaxID=92129;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99296737; PubMed=10366580;
RA Tidona C.A., Kurz H.W., Gelderblom H.R., Darai G.;
RT "Isolation and molecular characterization of a novel cytopathogenic
RT paramyxovirus from tree shrews."
RL Virology 258:425-434(1999).
CC -!- FUNCTION: THIS PROTEIN IS PROBABLY A COMPONENT OF THE ACTIVE
CC POLYMERASE. IT MAY FUNCTION IN TEMPLATE BINDING (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF079780; AAD28695.1; .
KW Transferase; RNA-directed RNA polymerase; Nucleocapsid;
KW Phosphorylation.
SQ SEQUENCE 527 AA; 57379 MW; BF21CE37F3224F06 CRC64;

alignment_scores:
Quality: 63.00 Length: 86
Ratio: 1.370 Gaps: 2
Percent Similarity: 53.488 Percent Identity: 24.419

alignment_block:
US-09-049-696-1 x RRPP_TPMV ..

Align seg 1/1 to: RRPP_TPMV from: 1 to: 527

1 GAATCACAAGGAGATGTACAGCAATGGGCCATTAAAGATTCTGTGTT 50
|||||:|||||: ||| |||:|||||: |||
355 GluIleLeuAsnArgLeuLysThrIleGlnSerLeuAsn 371
31 CATC.....TTGATTCTTCACCTTCAG 73
||| ||| |||
371 pSerIleLysArgIleGlnSerLysGlnGlyLeuAlaLeuSerThrLeuG 388
74 AAGGGCCCTGAGTAATTCATCTCAGCTGAACAACATGGCTATCAA 123
|||||:|||||: ||| |||: |||: |||
388 LuGlyLeuLeuSerValMetIleAlaIleProGlySerGlyAsnPro 404
124 GGCATTGTCTGTCGAATCGACCCCAATGTGCCA.....GA 158
||| ||| |||:|||||: |||
405 GlySerSerValGluIleAsnProAspLeuLysProMetLeuGlyArg 421
159 AGATCAACACACTCATTTCAACAATAAGGACATGTCACCCAGCATCTC 208
421 nLysAsnArgAlaLeuLysGluValSerAspGluLeuThrProAsnG 438
209 TGTATCTG 216
:: ||| ::
438 InPheLeu 440

seq_name: SwissProt_39:VG26_BPML5

seq_documentation_block:
ID VG26_BPML5 STANDARD; PRT; 836 AA.
AC G05233;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE MINOR TAIL PROTEIN GP26.
GN 26.
OS Mycobacteriophage L5.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
OX NCBI_TaxID=31757;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 1-10.
RX MEDLINE=93211282; PubMed=8459766;
RA Hatfull G.F., Sarkis G.J.;
RT "DNA sequence, structure and gene expression of mycobacteriophage L5:
RT a phage system for mycobacterial genetics."
RL Mol. Microbiol. 7:395-405(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z18946; CAA79402.1; .
DR PIR; S30971; S30971.
FT INIT MET .. 0
SQ SEQUENCE 836 AA; 86258 MW; 52E3040AA42BAD28 CRC64;

alignment_scores:
Quality: 63.00 Length: 72
Ratio: 1.575 Gaps: 1
Percent Similarity: 55.556 Percent Identity: 29.167

alignment_block:
US-09-049-696-1 x VG26_BPML5 ..

Align seg 1/1 to: VG26_BPML5 from: 1 to: 836

4 ATCACAGGAGATGTACAGCAATGGGCCATTAAAGATTCTGTGTTCAT 53
::

400 LeuThrGlyAlaLeuGlnAlaLeuGlyProIleLeuThrGlnValAlaTh 416
54 CTGTGATTCTTACAGAGGGCCCTGAGT..... 87
416 rLeuIleGlyThrLeuAsnThrAlaLeuGlnAlaLeuGlnProMetL 433
88 ..ANTTCATCTATTCAGTCAGACACATGGCTATGAGGCATGTCGTT 135
433 euProSerLeuMetGlnSerPheGlnIleSerAspValLeuValThr 449
136 GCAATCGACCCCAATGTCGAGAGATGAACACATCATTCACAAATAAA 185
450 SerLeuAlaProHisIleProAlaLeuAlaThrAlaLeuGlyGlnValAl 466
186 GGACATGGTGACCCAG 201
466 aGlyAlaValLeuGln 471

seq_name: SwissProt_39:AN11_HUMAN

seq_documentation_block:
ID AN11_HUMAN STANDARD: PRT: 84 AA.

AC Q9NYG5; O9P0R2;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ANAPHASE PROMOTING COMPLEX SUBUNIT 11 (HEPATOCYLLULAR CARCINOMA ASSOCIATED RING FINGER PROTEIN) (PROTEIN HSPC214).
GN ANAPC11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Chan A.H., Lee S.M.Y., Waye M.M.Y., Fung K.P., Lee C.Y., Tsui S.K.W.;
RT "Characterization of a human hepatocellular carcinoma associated ring finger protein.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20495367; PubMed=11042152;
RA Zhang Q.-H., Ye M., Wu X.-Y., Ren S.-X., Zhao M., Zhao C.-J., Fu G.,
RA Shen Y., Fan H.-Y., Lu G., Zhong M., Xu X.-R., Han Z.-G., Zhang J.-W.,
RA Tao J., Huang Q.-H., Zhou J., Hu G.-X., Gu J., Chen S.-J., Chen Z.;
RT "Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in CD34+ hematopoietic stem/progenitor cells.";
RT Genome Res. 10:1546-1560(2000).
RL
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A FRAMESHIFT IN POSITION 67.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF247565; AAF65816.1;
DR EMBL; AF151048; AAF36134.1; ALT_FRAME.
DR InterPro: IPR001841; Znf_ring.
DR SMART: SM00184; RING; 1.
KW Zinc-finger. 34 77 RING-TYPE.
FT ZN_FING
SQ SEQUENCE 84 AA; 9841 MW; EACBD5A54FDC11AE CRC64;

alignment_scores:

Quality: 62.50 Length: 35
Ratio: 3.289 Gaps: 2
Percent Similarity: 54.286 Percent Identity: 34.286
alignment_block:
US-09-049-696-1/rev x AN11_HUMAN ..
Align seg 1/1 to: AN11_HUMAN from: 1 to: 84
214 GATACAGACATGCGGTGACCATGTCCTTTATTTGTTCAATGACTGTT 165
||| :|||:|||||:|||||
36 AspCysLysValProGlyAspCysProLeuVal..... 47
164 TCATCTCTTGGCACATTTGGGTGATGCAACGACAAATCCCTTCATAGCC 115
|||||: :|||:|||||
48TrpGlyClnCysSer...HisCysPheHisMetH 58
114 ATTGT 110
58 ISCS 59

seq_name: SwissProt_39:AN11_MOUSE

seq_documentation_block:

ID AN11_MOUSE STANDARD: PRT: 84 AA.
AC Q9CPX9;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ANAPHASE PROMOTING COMPLEX SUBUNIT 11.
GN ANAPC11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kiehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wuzhuh-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RP "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AK003612; BAB22890.1;
DR EMBL; AK003244; BAB22663.1;
DR MGD; MGI:1913406; 111001119Rik.
DR InterPro: IPR001841; Znf_ring.

DT	01-APR-1990 (Rel. 14, Created)
DT	01-FEB-1994 (Rel. 28, Last sequence update)
DT	01-NOV-1997 (Rel. 35, Last annotation update)
DE	LIVER CARBOXYLESTERASE 2 (EC 3.1.1.1).
DE	Oryctolagus cuniculus (Rabbit).
DE	Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX	NCBI_TaxID=9986;
RP	[1]
RP	SEQUENCE.
RP	TISSUE=Liver;
RC	MEDLINE=89308686; PubMed=2745458;
RC	Ozols J.;
RA	"Isolation, properties, and the complete amino acid sequence of a
RT	second form of 60-kDa glycoprotein esterase. Orientation of the
RT	60-kDa proteins in the microsomal membrane.";
RT	J. Biol. Chem. 264:12533-12545(1989).
CC	-!- FUNCTION: INVOLVED IN THE DETOXIFICATION OF XENOBIOTICS AND IN
CC	THE ACTIVATION OF ESTER AND AMIDE PRODRUGS.
CC	-!- CATALYTIC ACTIVITY: A CARBOXYLIC ESTER + H(2)O = AN ALCOHOL
CC	+ A CARBOXYLIC ANION.
CC	-!- SUBUNIT: MONOMER.
CC	-!- SUBCELLULAR LOCATION: MICROSOMAL MEMBRANE, LUMEN OF ENDOPLASMIC
CC	RETICULUM.
CC	-!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
CC	PIR: A34329; A34329.
DR	HSSP: P21836; IMAH.
DR	InterPro: IPR002018; Carboxylesterase_B.
DR	InterPro: IPR000379; Est_lip_thioest_actise.
DR	Pfam: PF00135; Coesterase; 1.
DR	PROSITE: PS00122; CARBOXYLESTERASE_B.1; 1.
DR	PROSITE: PS00941; CARBOXYLESTERASE_B.2; 1.
DR	Glycoprotein; Hydrolase; Serine esterase; Endoplasmic reticulum.
KW	MOD_RES 1 1 PYRROLIDINE CARBOXYLIC ACID.
FT	ACT_SITE 201 201 BY SIMILARITY.
FT	ACT_SITE 430 430 BY SIMILARITY.
FT	DISULFID 69 96 BY SIMILARITY.
FT	DISULFID 253 264 BY SIMILARITY.
FT	SITE 529 532 PREVENT SECRETION FROM ER (POTENTIAL).
FT	CARBOHYD 249 249 N-LINKED (GLCNAC...).
SQ	SEQUENCE 532 AA; 59058 MW; C57DDDD76A13A9C52 CRC64;

alignment_scores:

Quality:	61.00	Length:	74
Ratio:	1.452	Gaps:	3
Percent Similarity:	56.757	Percent Identity:	27.027

alignment block:

US-09-049-696-1 X EST2_RABIT

Align seq 1/1 to: EST2_RABIT from: 1 to: 532

40 AGTTCTGTGTTCACTCTGATTCTTCACCTTCTAGAAGGG..... 78

277 ThrGlnValPheMetLeuIleProGlyValValAspGlyValPheLeuPr 293

79GCCCTGAGTAAT.....T 91

293 oArgHisProGluLeuLeuAlaLeuAlaAspPheGlnProValProS 310

92 CACTCATT CAGCTGAACAACAATGGCTATGAAGGCATTGTC..... 132

310 erlleleGlyleAsnAsnAspGluTyrGlyTrpilleleproLysLeu 326

133 ..GTTGCAATCGACCCCAATGTGCCAGAAAGATGAAACACTCATTTCAACA 179

327 LeuLeuAlaIleAspProGlnGluArqAspArqGlnAlaMetArqG1 343

180 AATAAGGACATGGTGACCCAG 201

343 u t l e M e t H i s G l n A l a T h r L y s 350

```

seq_name: SwissProt_39;WDR1_HUMAN
seq_documentation_block:
ID      WDR1_HUMAN          STANDARD;          PRT;       606 AA.
AC      Q75083; Q75131; Q9UQE0; Q9UG78;
DT      20-AUG-2001 (Rel. 40, Created)
DT      20-AUG-2001 (Rel. 40, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      WD-REPEAT PROTEIN 1 (ACTIN INTERACTING PROTEIN 1) (NORI-1).
GN      WDR1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Lymphoid;
RA      Abe Y., Nezu K., Ueda N.;
RL      Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
FN      [2]
RP      SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX      MEDLINE=99136857; PubMed=10036186;
RT      Adler H.-J., Winnicki R.S., Gong T.-W.L., Lomax M.I.;
RT      "A gene upregulated in the acoustically damaged chick basilar papilla
RT      encodes a novel WD40 repeat protein.";
RT      Genomics 56:59-69(1999).
KN      [3]
KN      SEQUENCE OF 58-606 FROM N.A.
RP      TISSUE=uterus;
RC      TISSUE=uterus;
RA      Wambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
RA      Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: INDUCES DISASSEMBLY OF ACTIN FILAMENTS IN CONJUNCTION
CC      WITH ADF/Cofilin FAMILY PROTEINS (BY SIMILARITY).
CC      -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC      SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC      -1- SIMILARITY: CONTAINS 11 WD REPEATS (TRP-ASP DOMAINS).
CC      -1- SIMILARITY: BELONGS TO THE AIP1 FAMILY OF WD-REPEAT PROTEINS.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outpost -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to licensel@isb-sib.ch).
CC      -----
EMBL: AB010427; BAA31855.2; -
EMBL: AF020056; AAD05044.1; -
EMBL: AF020260; CAD05045.1; -
EMBL: AL050108; AAB03276.1; -
EMBL: MIM: 604734; -.
DR      InterPro: IPR001680; WD40.
DR      Pfam: PF00400; WD40; 10.
DR      PRINTS: PR00320; GPROTEINBRPT.
DR      SMART: SM00320; WD40; 11.
DR      PROSITE: PS00678; WD_REPEATS_1; 1.
DR      PROSITE: PS50082; WD_REPEATS_2; 5.
DR      PROSITE: PS50294; WD_REPEATS_REGION; 1.
KW      Actin-binding; Cytoskeleton; Repeat; Alternative splicing.
FT      REPEAT             56..95
FT              WD 1.
FT      REPEAT             100..143
FT              WD 2.
FT      REPEAT             144..184
FT              WD 3.
FT      REPEAT             187..226
FT              WD 4.
FT      REPEAT             232..271
FT              WD 5.
FT      REPEAT             318..359
FT              WD 6.
FT      REPEAT             363..401
FT              WD 7.
FT      REPEAT             443..482
FT              WD 8.
FT      REPEAT             487..526
FT              WD 9.
FT      REPEAT             530..569
FT              WD 10.
FT      REPEAT             574..605
FT              WD 11.
FT      VARSPPLIC         533..606
FT      AKIVCLVAVSPDNEHFASGGMDMMVVYWTLLSDPETRYKIODA
FT      HRIHLHVSLSLAWLDEHTLVTTSHDSVKVENTIV -> EK
FT      (IN SHORT ISOFORM)

```


This Page Blank (uspto)

OM of: US-09-049-696-1 to: PIR_68.* out_format : pfs

Date: Mar 30, 2002 2:25 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame_n2p.model -DBV=xl
-Q/cgnt2/uspto_spool/us09049696/runat_28032002_145238_2031/app_query.fasta_1.12579
-DB=PIR_68 -OFMT=fastan -SUFFIX=tr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -EGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTPMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09049696_@CGN1_1_367 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-049-696-1
Query length: 223
Database: PIR_68.*
Database sequences: 219241
Database length: 76174552
Search time (sec): 470.790000

score_list:	Sequence	Strd Orig	ZScore	EScore	Len	Documentation
pir2:JG0168	+ 257.00	567.77	1.4e-24	913	1	gob-5 protein - mouse
pir2:T02205	+ 181.50	393.85	6.6e-15	905	1	Lu-ECAM-1 protein - bovine
pir2:T10997	- 73.00	157.07	0.5341	174	1	NADH dehydrogenase (ubiquinone)
pir2:T44004	+ 66.00	138.66	4.27	231	1	hypothetical protein U44 [impod
pir2:A96750	+ 66.00	126.94	4.46	993	1	hypothetical protein F29p22.25
pir2:T18418	+ 65.50	122.93	5.23	1417	1	hypothetical protein C0120w
pir2:A42249	- 64.00	127.74	7.89	507	1	serine-type carboxypeptidase (P
pir2:T26252	+ 63.50	125.24	9.19	599	1	hypothetical protein W06H3.3
pir2:S30971	+ 63.00	121.40	10.77	837	1	gene 26 protein - Mycobacterium
pir1:HM1VSA	- 61.50	121.09	16.58	566	1	hemagglutinin precursor - influ
pir2:A18160	- 61.50	118.60	16.73	772	1	tetracycline resistance element
pir2:S58132	+ 61.00	122.22	19.05	426	1	Sls1 protein precursor - yeast
pir2:T34329	+ 61.00	120.44	19.18	532	1	60K esterase (EC 3.1.1.-) isofc
pir2:T13152	+ 61.00	119.39	19.26	606	1	WDR1 protein - human
pir2:C85017	+ 61.00	114.48	19.62	1117	1	hypothetical protein A74g01400
pir2:D82908	+ 60.50	127.73	21.53	186	1	hypothetical protein U0309 [imp
pir2:A86615	+ 60.50	122.49	21.96	357	1	CT839 hypothetical protein [imp
pir2:D72009	+ 60.50	122.49	21.96	357	1	conserved hypothetical protein
pir2:A55173	+ 60.50	115.40	22.57	863	1	cf-9 protein precursor - tomato
pir2:E71508	+ 60.00	126.41	24.97	190	1	probable methylase - Chlamydia
pir2:S67166	+ 60.00	118.73	25.72	494	1	PAC1 protein - yeast (Saccharom
pir2:E70234	+ 59.50	117.67	29.80	488	1	conserved hypothetical protein
pir2:JED0346	+ 59.50	116.61	29.92	557	1	high-affinity carnitine transpo
pir2:T73484	+ 59.50	115.79	30.02	617	1	hypothetical protein - Deinococ
pir2:T33698	+ 59.00	119.67	34.13	330	1	protein tyrosine phosphatase -
pir2:F82888	- 59.00	117.45	34.42	435	1	GTP-binding protein U0461 [impc
pir2:S18446	+ 59.00	116.81	34.51	471	1	variant surface glycoprotein in
pir2:S18423	+ 59.00	114.91	34.76	596	1	hypothetical protein yj9d [impc
pir2:D83870	+ 59.00	112.92	35.02	764	1	ATP-dependent RNA helicase BHL7
pir2:F86272	+ 58.50	124.43	38.69	158	1	initiation factor 5A-4 [importe
pir2:H81261	+ 58.50	124.03	38.75	166	1	hypothetical protein Cj1650 [im
pir2:T42365	+ 58.50	120.71	39.24	251	1	uroporphyrin-III C-methyltransf
pir1:RWMSC8	+ 58.50	120.22	39.31	267	1	T-cell receptor alpha chain - m
pir2:T13317	+ 58.50	119.36	39.44	297	1	hypothetical protein 28 - Strept
pir2:S48370	+ 58.50	115.69	40.00	469	1	probable phosphoglycerate dehyd
pir2:T41114	+ 58.50	115.24	40.07	496	1	uroporphyrin methyltransferase
pir2:D64244	+ 58.50	114.89	40.12	518	1	H+-transporting ATP synthase (F
pir1:HM1VTV	- 58.50	114.26	40.22	560	1	hemagglutinin precursor - influ
pir2:D70199	+ 58.50	110.80	40.75	862	1	DNA mismatch repair protein (mu
pir2:A82272	+ 58.50	107.51	41.27	1297	1	phosphoribosylformylglycinamid
pir1:S43793	+ 58.50	106.39	41.44	1491	1	copper-transporting ATPase (ED

pir2:H55110 + 58.00 121.88 45.09 188 1 hypothetical 20.0 kD protein
pir2:H55983 + 58.00 121.88 45.09 188 1 hypothetical protein Yrbi 11
pir2:Y51204 + 58.00 117.68 45.82 317 1 hypothetical protein B7M4.11
pir2:A71903 + 58.00 116.96 45.94 347 1 D-alanine--D-alanine ligase
seq_name: pir2:JG0168
seq_documentation_block:
gob-5 protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-Mar-2000
C:Accession: JG0158
E:Komiya, T.; Taniqawa, Y.; Hirohashi, S.
Biochem. Biophys. Res. Commun. 255, 347-351, 1999
A:Title: Cloning and identification of the gene gob-5, which is expressed in intestine
A:Reference number: JG0168; MUID:99160866
A:Accession: JG0168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-913 <ROM>
A:Cross-references: DDBJ:AB016592

alignment_scores:
Quality: 257.00 Length: 66
Ratio: 4.213 Gaps: 0
Percent Similarity: 92.424 Percent Identity: 80.303

alignment_block:
US-09-049-696-1 x JG0168

Align seg 1/1 to: JG0168 from: 1 to: 913

```
25 ATGGGGCCATTAAAGAGTTCTGTGTTTCATCTTGTATCTTCACCTCTCTAGA 74
   ||| :||| ||||| :||| :||| :||| :||| :||| :||| :||| :|||
   1 MetGluSerLeuLysSerProValPheLeuLeuIleLeuHisLeuLeuG 17

75 AGGGGCCCTGAGTAATTCATCTACTCATTGAGCTGACACATGGCTATGAAG 124
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
   17 uGlyValLeuSerGluSerLeuIleGlnLeuAsnAsnGlyTyrGluG 34

125 GCATCTGCTGCTCAATCGACCCCAATGTCGACAGACATGAACACATCATT 174
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
   34 LylleValIleAlaIleLeuPheLeuAspValProGluAspGluAlaLeu 50

175 CAACAAATAAGACATGCTGACCCAGGATCTCTCTATCTGTTTCAA 222
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
   51 GlnHisIleLysAspMetValThrGlnAlaSerProTyrLeuPheGlu 66
```

seq_name: pir2:T02205

seq_documentation_block:
Lu-ECAM-1 protein - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 01-Dec-2000
C:Accession: T02205; T02152; T02171
E:Elhag, R.C.; Widom, J.; Gruber, A.D.; Abdel-Ghany, M.; Levine, R.; Goodwin, A.; Paul
submitted to the EMBL Data Library, April 1997
A:Description: Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial
A:Reference number: 214590
A:Accession: T02205
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-905 <ELH>
A:Cross-references: EMBL:AF001261; NID:g2623762; PIDN:AAB86529.1; PID:g2623763
A:Experimental source: lung
A:Accession: T02152
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-793 <SGSP> 'M', 862, 'RF', 865-866, 'O', 868, 'AKVLELO', 876, 'OHQ', 880, 'FQ', <
A:Cross-references: EMBL:AF001263; NID:g2623766; PIDN:AAB86531.1; PID:g2623767
A:Experimental source: lung
A:Accession: T02171

```
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-792, 'ES', <EL3>
A:Cross-references: EMBL:AF001262; NID:g2623764; PIDN:AAB86530.1; PID:g2623765
A:Experimental source: lung

alignment_scores:
  Quality: 181.50      Length: 58
  Ratio: 3.490         Gaps: 1
  Percent Similarity: 89.655      Percent Identity: 60.345

alignment_block:
US-09-049-696-1 x T02205

Align seg 1/1 to: T02205 from: 1 to: 905
46 GTGTTTCATCTTGATCTTACCTTCTAGAGGGCCCTGAGTAATTCAC 95
8 IleLeuPheLeuThrLeuHisLeuLeuProGly...MetIysSerSerMe 23
96 CATTCAGCTGCAACCAATCGCTATGAGGCAATGCTCTTGAATCAACC 145
23 tValAsnLeuIleAsnAsnGlyTyrAspGlyIleValIleAlaIleAsn 40
146 CCATGTCGCCAGAGATCAACACTCATTCAACAATAAAGGACATGGTG 195
40 roSerValProGluAspGluLeuLeuGluAsnIleLysGluMetVal 56
196 ACCCAGGCATCTCTGCTATCTCTTT 219
57 ThrGluAlaSerThrTyrLeuPhe 64

seq_name: pir2:T10997

seq_documentation_block:
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 6 - Chrysemys picta mitochondrion
C:Species: mitochondrion Chrysemys picta
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 11-Jan-2000
C:Accession: T10997
R:Sorenson, M.D.; Dimcheff, D.E.; Ast, J.C.; Yuri, T.; Mindell, D.P.
submitted to the EMBL Data Library, September 1998
A:Description: Complete mitochondrial DNA sequences for five birds and a turtle.
A:Reference number: 217239
A:Accession: T10997
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-174 <SOR>
A:Cross-references: EMBL:AF069423; NID:g4530187; PID:g4530199; PIDN:AAD21936.1
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 6
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

alignment_scores:
  Quality: 73.00      Length: 28
  Ratio: 3.318         Gaps: 0
  Percent Similarity: 78.571      Percent Identity: 32.143

alignment_block:
US-09-049-696-1/rev x T10997

Align seg 1/1 to: T10997 from: 1 to: 174
192 CATGCTCCTTTATTTGTTGAATGAGTGTTCATCTCTGCGACATGGGGT 143
3 TyrPheMetPheLeuPheGlyPheCysPheValPheTrpMetValGlyVa 19
142 CGATTGCAACGACCAATGCTTCATAGCAATGGTT 109
19 lSerCysAsnProSerProTyrTyrGlyValLeu 30
```

```
seq_name: pir2:T44004

seq_documentation_block:
hypothetical protein U44 [imported] - human herpesvirus 6
C:Species: human herpesvirus 6
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T44004; T44191
R:Isegawa, Y.; Mukai, T.; Nakano, K.; Kagawa, M.; Chen, J.; Mori, Y.; Sunagawa, T.; K
J. Virol. 73, 8053-8063, 1999
A:Title: Comparison of the complete DNA sequences of human herpesvirus 6 variants A a
A:Reference number: Z22732; MUID:99412319
A:Accession: T44004
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-231, <ISE>
A:Cross-references: EMBL:AB021506; NID:g495977; PIDN:BAA78265.1; PID:g4996032
R:Experimental source: strain HST; pop. variant B
R:Domiguez, G.; Dambaugh, T.R.; Stamey, F.R.; Dewhurst, S.; Inoue, N.; Pellett, P.E.
J. Virol. 73, 8040-8052, 1999
A:Title: Human herpesvirus 6B genome sequence: coding content and comparison with hum
A:Reference number: Z22734; MUID:99412318
A:Accession: T44191
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-231 <DOM>
A:Cross-references: EMBL:AF157706; PIDN:AAB06342.1
A:Experimental source: strain Z29; variant B
C:Genetics:
A:Gene: U44

alignment_scores:
  Quality: 66.00      Length: 39
  Ratio: 2.276         Gaps: 0
  Percent Similarity: 74.359      Percent Identity: 33.333

alignment_block:
US-09-049-696-1 x T44004

Align seg 1/1 to: T44004 from: 1 to: 231
88 AATTCATCTATTAGCTGAACAATGGCTATGAAGGCAATTCGTTGC 137
158 AsnSerLeuIleGluIleAspGlyGluAsnSerIysArgLeuValG1 174
138 AATCGACCCCAATGTCGCAGAGATGAACACTCATTCACAAATAAGG 187
174 uLeuAspProIleLeuHisGluGluThrGlyLeuTyrGlnAlaLeuProA 191
188 ACATGTTGACCCAGGCA 204
191 snValValThrGluAla 196

seq_name: pir2:A96750

seq_documentation_block:
hypothetical protein F28P22.25 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A96750
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, W.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: A96750
```

A;Cross-references: GB:AE005173; NID:q6648172; PIDN:AAF21172.1; GSPDB:GN00141

204 ATCTCTGTAT 213
:::||||:
.56' pGlutLeuHis 69

```
seq_documentation_block:
  serine-type carboxypeptidase (EC 3.4.16.-) sxa2 - fission yeast (Schizosaccharomyces
  C:Species: Schizosaccharomyces pombe
  C:Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 20-Jun-2000
  C:Accession: B42249; T37564
```

serine_type calixosaccharomycase (EC 3.4.16.1) skaz - fission yeast (schizosaccharomycase)
C:Species: Schizosaccharomycase pombe
C:Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 20-Jun-2000
C:Accession: B42249; T37564

C/Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 20-Jun-2000
C/Accession: B42249; T37564

K. IMAI, I.; IANAMOTO, M.
Mol. Cell. Biol. 12: 1827-1834. 1992

A:Title: Schizosaccharomyces pombe sxa1(+) and sxa2(+) encode putative proteases involved in the regulation of the cell cycle

A;Accession: B42249
A;Status: preliminary

A;Molecule type: DNA
A;Residues: 1-507 <IMA>

A/Cross-references: GB:D10199; NTD:g218559; PIDN:BAA01047.1; PID:g218560
R;Wood, V.; Barrell, B.G.; Rajandream, M.A.; Harris, D.; Seeger, K.

submitted to the EMBL Data Library, February 1999
A; Reference number: Z21725

A;Accession: T37564
A;Status: preliminary; translated from GB/EMBL/DBJ

A; Molecule type: DNA
A; Residues: 1-507 <WOO>

A; Cross-references: EMBL:ALU35439; PIDN: CAB36509.1; GSPDB: GN00066; SPDB: SPAC1296.036
A; Experimental source: strain 972h-; cosmid cl296

C;Genetics: A;Gene: *sxa2*
A:Map position: 1

C;Superfamily: serine carboxypeptidase
C;Keywords: hydrolase, serine carboxypeptidase

.....

```
alignment_scores:
  alignment_scores:
    Quality: 64.00
    Length: 38
```

Ratio:	2.462	Gaps:	1
Percent Similarity:	68.421	Percent Identity:	34.211

alignment_block:

US-03-049-696-1/rev x B42249 ..

Align seq 1/1 to: B42249 from: 1 to: 507

272 1TTCAACAGATACAGAGATGC.....CTGGGTCCACCATGTCCT 185
||| :::: ||| |::| :
273 PhageCotE1-069T-EcoRI-EcoRV-BstXI-PfuI-SbfI-HpaII-XbaI-KpnI-NotI-DraIII-220
||||: ||| :|::| :|::| :|::| :

322. *measlin* *seil* *lle* *ser* *gly* *cys* *asp* *preu* *r* *ser* *leu* *ser* *asp* *phen* *leu* *le* 338

184 TATATTGTTGAATAGAGTGTTCATCTCTCGGCACATTTGGGGTGCGATTGCC 135
339 GTTATTCATCCTAAAGSerGueValTLebMethylenValSotLysArgGue 355

[illegible]

355 **erpPheAsnGluTyr** 359

```
seq name: pir2:"26252
```

```
seq_documentation_block;
```

hypothetical protein W06H3.3 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*

```
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
```

```

417 rLeuIleGlyThrThrLeuAsnThrAlaLeuGlnAlaLeuGlnPrometL 434
88 ..AATTCACCTCATTCAGCTGAACAACAATGGGTATGAAGCATTGTGCTT 135
      |||||:|||||:|||||:|||||:|||||:
434 euProSerLeuMetGlnSerPheGlnGlnIleSerAspValLeuValThr 450
      :|||:
136 GCATTCAGCCCCCAATGTGCCAGAGATGAACAACACTCATTCACAACAATAA 185
      |||||:|||||:|||||:|||||:|||||:
451 SerLeuAlaProHisIleProAlaLeuAlaThrAlaLeuGlyGlnValAl 467
186 GGACATGGTGACCCAG 201
      ||| |||
467 aGlyAlaValLeuGln 472
seq.name: p1r1:HWYSA

```

hemagglutinin precursor - influenza A virus (strain A/shearwater/Australia/1/72 [H6N2])
N:Contains: hemagglutinin HA1; hemagglutinin HA2
C:Species: Influenza A virus
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 16-Jun-2000
C:Accession: E39987; B33157
R:Nobunawa, E.; Aoyama, T.; Kato, H.; Suzuki, Y.; Tateno, Y.; Nakajima, K.
Virology 182, 475-485, 1991
A:Title: Comparison of complete amino acid sequences and receptor-binding properties
A:Reference number: A39987; MUID:91220697
A:Accession: E39987
A:Molecule type: genomic RNA
A:Residues: 1-566 <NOB>
A:Cross-references: GB:D90303; NID:g221315; PIDN:BAA14333.1; PID:g221316
C:Genetics:
A:Map position: segment 4
C:Superfamily: Influenza virus hemagglutinin
C:Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond; tr

```

F:17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F:345-566/Product: hemagglutinin HA2 #status predicted <HA2>
F:539-555/Domain: transmembrane #status predicted <TM1>
F:267-301/182,306,498/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:553,562,563/Binding site: palmitate (Cys) (covalent) #status predicted

alignment_scores:
  Quality: 61.50      Length: 60
  Ratio: 1.922      Gaps: 2
  Percent Similarity: 53.333      Percent Identity: 26.667

alignment_block:
  US-09-049-696-1/rev x HMI_VSA ..

Align seg 1/1 to: HMI_VSA from: 1 to: 566

222 TTCAACAGATACAGAGATGCGTGGGTCCACCATGTCCTTTATTGTTGAA 173
      :::::::::::::: ||| :::::::::::::: ||| ||
453 TyrGluuArgValIysSerGlnLeuArgAspAsnAlaMetIleLeuCllyas 479

```

```

479 nGlycyspHeGluPheTrpHis.....LysCysAspAspGluCysM 493
122 TCATAGCCATTTGTTGTTTCAGCTGAATGACGTAATCTACTCAGGGCCCTTC 73
493 eL.....G.....GluSerValLysAsnGlyThrTyr 501
72 TAGAAGGTGAAGAATCAAGATGAACACAGA 43
502 AspTyrProLysTyrGlnAspGluSerLys 511
seq_name: p1r2:A41860
seq_documentation_block:
tetracycline resistance element regulator RteA - Bacteroides th
C:Species: Bacteroides spp.

```

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 18-Aug-2000

C:Accession: A41860
R:Stevens, A.M.; Sanders, J.M.; Shoemaker, N.B.; Salyers, A.A.
J. Bacteriol. 174, 2935-2942, 1992
A:Title: Genes involved in production of plasmidlike forms by a Bacteroides conjugal chr
A:Reference number: A41860; MUID:92234952
A:Accession: A41860
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-772 <STP>
A:Cross-references: GB:M81439; GB:M81881; NID:g143966; PIDN:AAA22920.1; PID:g143968
A>Note: sequence extracted from NCBI backbone (NCBIP:97351)
A:Note: the source species is not identified
C:Superfamily: response regulator homology
C:Keywords: phosphohistidine; phosphoprotein
F:309/Binding site: phosphate (His) (covalent) #status predicted
F:588/Binding site: phosphate (Asp) (covalent) #status predicted

alignment_scores:

Quality: 61.50 Length: 64
Ratio: 1.577 Gaps: 3
Percent Similarity: 60.938 Percent Identity: 31.250

alignment_block:

US-09-049-696-1 x A41860 ..

Align seg 1/1 to: A41860 from: 1 to: 772

37 AAGAGTCTGTGTTGATCTTGTATCTTACCTCTAGAA..... 75
|||||... |||
506 LysGlySerArgPheThrValGluIleSerMetGlnGluAlaGluGlu 522
76 ...GGGGCCCTGAGTAATTCACCTACCTGACCTGAACAACAAATGGCTATG 121
|||||... |||
522 nLeuGlyTyrThrSerAsnThrProValTyrHisAsnLysPheHis 539
122 AAGGCATTCGTTGTAATCGACCCCAATGTGCCAGAGATGAACACTC 171
:: |||||... |||
539 sp.....ValValAlaIleAsp.....AsnAspGluValLeu 549
172 ATTCAACAATAAGACATGTGGACCCAGGCATCTCTGTAT 213
::: |||
550 LeuLeuMetLeuLysGluMetTyrSerGlnGluGlyIleHis 563

seq_name: pir2:S58132

seq_documentation_block:

Sls1 protein precursor - yeast (Yarrowia lipolytica)
C:Species: Yarrowia lipolytica, Candida lipolytica
C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 09-Sep-1997
C:Accession: S58132
R:Boisrame, A.; Beckerich, J.; Gaillardin, C.
submitted to the EMBL Data Library, July 1995
A:Description: Sls1p, an endoplasmic reticulum component, is involved in the protein tra
A:Reference number: S58132
A:Accession: S58132
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-426 <BOI>
A:Cross-references: EMBL:250154; NID:g1052827; PID:g1052828

alignment_scores:

Quality: 62.00 Length: 56
Ratio: 1.550 Gaps: 2
Percent Similarity: 71.429 Percent Identity: 33.929

alignment_block:

US-09-049-696-1 x S58132 ..

Align seg 1/1 to: S58132 from: 1 to: 426

61 CTTCACTT.....CTAGAAGGGCCCTGAGTAATTCACCTATTCA 101
|||||... |||
59 LeuHisValArgMetAspTyrGluAsnGlyValLysGluAlaLysIleAs 75
102 GCTGAACAACAATGGCTATTCAGAGCAATTCGTTGCAATCGACCCCAATG 151
: |||||... |||
75,nAspProAsnGluGluValGluGlyValAlaValAlaValGlyGluGluV 92
152 TGCAGAAGATGAACACACCTTCATCAACAATAAAGGACAT....GCTGAC 197
|||||... |||
92,alProGluGlyGluValValIleGluAspLeuThrGluGluAsnGlyAsp 108
198 CCAGGCATCTCTGTATCT 215
: |||||... |||
109 GluGlyIleSerAlaAsn 114
seq_name: pir2:A34329

seq_documentation_block:

60K esterase (EC 3.1.1.-) isoform 2 - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 21-Aug-1998
C:Accession: A34329
R:Ozols, J.
J. Biol. Chem. 264, 12533-12545, 1989
A:Title: Isolation, properties, and the complete amino acid sequence of a second form
A:Reference number: A34329; MUID:89308686
A:Accession: A34329
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-532 <OZO>
C:Superfamily: cholinesterase; cholinesterase homology
C:Keywords: carboxylic ester hydrolase
F:32-517/Domain: cholinesterase homology <CHE>
F:201,430/Active site: Ser, His #status predicted

alignment_scores:

Quality: 61.00 Length: 74
Ratio: 1.452 Gaps: 3
Percent Similarity: 56.757 Percent Identity: 27.027

alignment_block:

US-09-049-696-1 x A34329 ..

Align seg 1/1 to: A34329 from: 1 to: 532

40 AGTTCTGTTTCATCTTGTATCTTACCTCTAGAGG..... 78
: |||||... |||
277 ThrGlnValPheMetLeuIleProGlyValValAspGlyValPheLeuPr 293
79GCCCTGAGTAAT.....T 91
|||||...
293 oArgHisProGluGluLeuAlaLeuAlaAspPheGlnProValProS 310
92 CACTCATTCAGCTGACACACATGGCTATGAGGCATTGTC..... 132
: |||||... |||
310 erIleIleGlyIleAsnAsnAspGlyTyrGlyTrpIleIleProLysLeu 326
133 ...GTTGCATCGACCCCAATGTGCCAGAGATGAACACTCATTCACA 179
: |||||... |||
327 LeuLeuAlaIleAspProGlnGluGluArgAspArgGlnAlaMetArgG 343
AATAAAGGACATGGTGACCCAG 201
|||
343 ulleMetHisGlnAlaThrLys 350

seq_name: pir2:T13152

seq_documentation_block:

WDRL protein - human
C:Species: Homo sapiens (man)
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C:Accession: T13152; T08756; T14754
R:Adler, H.J.; Winnicki, R.S.; Gong, T.W.; Lomax, M.I.
Genomics 56, 59-69, 1999
A:Title: A gene upregulated in the acoustically damaged chick basilar papilla encodes a
A:Reference number: Z17609; MUID:99156857
A:Accession: T13152
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-606 <ADL>
A:Cross-references: EMBL:AF020056; NID:g3420178; PID:g3420179; PIDN:AAD05044.1
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16469
A:Accession: T08756
A:Molecule type: mRNA
A:Residues: 58-184, 'V', 186-606 <WAM>
A:Cross-references: EMBL:AL050108
A:Experimental source: adult uterus; clone DKFzP586I2219
R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, August 1999
A:Reference number: Z18180
A:Accession: T14754
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 'R', 139-140, 376-606 <KOE>
A:Cross-references: EMBL:AL110208
A:Experimental source: adult uterus; clone DKFzP586P1922
C:Genetics:
A:Map position: 4
A:Note: DKFzP586I2219.1; DKFzP586P1922.1
C:Superfamily: unassigned WD repeat proteins; WD repeat homology

alignment_scores:
Quality: 61.00 Length: 49
Ratio: 1.906 Gaps: 0
Percent Similarity: 65.306 Percent Identity: 28.571

alignment_block:

US-09-049-696-1 x T13152 ..

Align seg 1/1 to: T13152 from: 1 to: 606

```
1 GAATCAGGAGATGACAGCAATGGGCGCATTAAGATTCTGTGTT 50
      :::::  :::::  :::::  :::::  :::::  :::::
407 AspValGlnProLysCysValAlaValGlyProGlyGlyTyrAlaValVa 423
      ::  |||  :::::  :::::  :::::  :::::  :::::
51 CATCTTGATTCTTCACTTCTAGAGGGCCCTGAGTAACTCACTATTC 100
      ::  |||  :::::  :::::  :::::  :::::  :::::
423 lValCysIleGlyGlnIleValLeuLeuLysAspGlnargLysCysPheS 440
      :::::  :::::  :::::  :::::  :::::  :::::
101 AGCTGACACAAATGGCTATGCAAGCATTTGCTTGCATTCACGCC 147
      :::::  :::::  :::::  :::::  :::::  :::::
440 erIleAspAsnProGlyTyrGluProGluValValAlaValHisPro 455
```

seq_name: pir2:C85018

seq_documentation_block:

hypothetical protein AT4g01400 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: C85018
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488
A:Accession: C85018
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1117 <STO>
A:Cross-references: GB:NC_001268; NID:g7267637; PIDN:CAB80949.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g01400

A:Map position: 4
alignment_scores:
Quality: 61.00 Length: 81
Ratio: 1.525 Gaps: 3
Percent Similarity: 49.383 Percent Identity: 29.630
alignment_block:
US-09-049-696-1 x C85018 ..
Align seg 1/1 to: C85018 from: 1 to: 1117
1 GAATCAGGAGATGACAGCAATGGGCGCA..... 33
||| ::::: ::::: ::::: ::::: ::::: :::::
901 GluIleGluGluGlnCysThrGluValPheProAlaProAlaAspArgG1 917
34TTTAAGAGTCTGTGTTTCATCTTTCACCTTCTAGAGGGG 79
||| ::::: ::::: ::::: ::::: ::::: :::::
917 uArgIleLysSerCysLeuSerGluLeu.....GlyG 928
30 CCCTGAGTAAATTCATCTCATTCAGCTGAACAAATGGCTATCAAGGCATT 129
||| ::::: ::::: ::::: ::::: ::::: :::::
928 luLeuSerSerThrPheLysGlnLeuLeuAsnSerGlyMetGluGlnLeu 944
130 GTCGTTGCAATCGACCCCAATGTGCCAGAGATGAAACACTCATTCACA 179
||| ::::: ::::: ::::: ::::: ::::: :::::
945 ValAlaThrValThrProArg.....IleArgPr 954
180 AATAAGGACATGGTGACCCAGGCATCTCTATCTGTTTCAA 222
::: ||| ::::: ::::: ::::: ::::: :::::
954 oValLeuAspThrValAlaThrIleSerTyrGluLeuThrGlu 968

alignment_scores:
Quality: 61.50 Length: 90
Ratio: 1.281 Gaps: 6
Percent Similarity: 53.333 Percent Identity: 28.889

alignment_block:
US-09-049-696-1 x US-09-082-593-10 ..
Align seg 1/1 to: US-09-082-593-10 from: 1 to: 270

```
15 ATGTACACCAATGGGCCATTACAG.....TTCTGTGTCATCTT... 56
70 IIEPESerASnglygluLysGluGluGlyArgPheThrIleHisLeuAs 86
57 .....GATTCTTCACCTCTCTA 72
86 nLysAlaSerLeuHisPheSerLeuHisIleArgAspSerGlnProSer. 102
73 GAAGGGCCCTGAGTAATCTACTCATTCAGCTG...ACAACAATGGCTA 119
103 AspSerAlaLeuTyLeuCyAlaValThrLeuTyrglyGlySerGlyAs 119
120 TGAA.....GCCATTGTCTTCAATCGACCCCAATGTC 154
119 nLysLeuIlePheGlyThrGlyThrLeuLeuSerValLysProAsnIleG 136
155 CAGAAGATGAACACTCATTCACAAATAAAGAGCATGGTGACCCAG... 201
136 InAsnProGluProAlaValTyrglnLeuLysAspProArgSerGlnAsp 152
202 GCATCTCTGTATCTGTTT 219
153 SerThrLeuCyLysLeuPhe 158
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-883-534-3

seq_documentation_block:
Sequence 3, Application US/08883534
Patent No. 5846777
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: TWO NEW WD-40 PROTEINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/883,534
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0332 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166

TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 606 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: NEUTGMT01
CLONE: 1221143
US-08-883-534-3

alignment_scores:
Quality: 61.00 Length: 49
Ratio: 1.906 Gaps: 0
Percent Similarity: 65.306 Percent Identity: 28.571

alignment_block:
US-09-049-696-1 x US-08-883-534-3 ..
Align seg 1/1 to: US-08-883-534-3 from: 1 to: 606

```
1 GAATCACAGGAGATGTACACCAATGGGGCCATTAAAGACTTCGTGCTT 50
407 AspValGlnProLysCysValAlaValGlyProGlyGlyTyrAlaValVa 423
51 CATCTTGATTCTTCACCTTCTAGAGGGCCCTGAGTAATTCACCTCATTC 100
423 lValCysIleGlyGlnIleValLeuLeuLysAspGlnArgLysCysPheS 440
101 AGCTGAACACCAATGGGTATGAAGGCATTTGCTGTGCAATCGACCCC 147
440 erIleAspAsnProGlyTyrgluProGluValValAlaValHisPro 455
```

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-09-204-764-3

seq_documentation_block:
Sequence 3, Application US/09204764
Patent No. 6025464
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: TWO NEW WD-40 PROTEINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/204,764
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/883,534
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0332 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:

ATTORNEY/AGENT INFORMATION:
NAME: MISFOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-0A-185-432-17

alignment_scores:		Length:	69
	Quality:	Gaps:	4
	Ratio:	Percent Identity:	28.986
Percent Similarity:	49.275		

alignment_block:
US-02-049-696-1/rev x US-08-185-432-17

Align seg 1/1 to: US-08-185-432-17 from: 1 to: 2556

198 GGHCACCAAGTCCTTTATTTGTTTGAAGTAGCGTTTTCATC..... 160
|||||
597 GlyHisHisCysGluThrAsnIleAsnGluCysSerSerGlnProCysAr 613
|||
159..TTCTGGCACATTGGGTGCATTCACAGGACAATGCCCTTCATAGCCATTG 111
||||| ||| |||||||
613 gtleutrpGlyThrCysGlnAsp...ProAspAsnAlatyrLeuCysPheC 629
|||||
110 TTGTTTCACGTGAATGAGTGAACTACTCAGGCCGCCCTTCTAGAAGGTGAAG 61
||| |||||
629 ysLeuLys.....GlyThrThrGlyProAsnCysGluIleAsn 641
|||
60 ATTCAGAGTGAACACAGAACTCTTTAAATGGCCCCCATTCCTGTACATCTCC 11
:|||
642 LeuAspAsp.....CysAlaSerSerPr 649

10 CTGTCAT,4

```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-983-607-2
seq_documentation_block:
: Sequence 26, Application US/098933607
: Patent No. 6140470
: GENERAL INFORMATION:
: APPLICANT: Alan Garen
: APPLICANT: Xiaohong Cai
: TITLE OF INVENTION: Human Anti-Tumor Monoclonal Anti-
: bodies
: NUMBER OF SEQUENCES: 51
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Department of Molecular Biophysics
: ADDRESSEE: and Biochemistry, Yale University
: STREET: 266 Whitney Avenue
: CITY: New Haven
: STATE: Connecticut
: COUNTRY: United States of America
: Zip: 06520-8114
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" 1.44 Mb diskette
: COMPUTER: IBM PC
: OPERATING SYSTEM: MS DOS
: SOFTWARE: word processing
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/983.607

```



```

; ZIP: 94304
; COMPUTER AVAILABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/224,917
; FILING DATE: 08-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Peries, Rohan
; REGISTRATION NUMBER: 35,752
; REFERENCE/DOCKET NUMBER: 28060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)-852-1698
; TELEFAX: (415)-496-3529
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 693 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-224-917-11

alignment_scores:
      Quality: 57.50      Length: 57
      Ratio: 1.643      Gaps: 1
      Percent Similarity: 61.404      Percent Identity: 24.561

alignment_block:
US-09-049-696-1 x US-08-224-917-11 ..

Align seg 1/1 to: US-08-224-917-11 from: 1 to: 693

79 GCCTGTGACATAATTCACATCATTCAGCTCAACCAACAATGGC..... 117
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2 AlaLeuCysAsnGlyAspSerLysLeuGluAsnAlaGlyAspLeuLys 18
118 .....TATGAAGCATTTGCTTGCATCGACCCCAATG 151
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
18 sAspGlyHisHisHisTyrGluGlyAlaValValIleLeuAspAlaGlyA 35
152 TCCCAGAAGATGAACACTCATTCACAAATAAAGACATGGTGACCCAG 201
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
35 laGlnTyrGlyLysValIleAspArgValArgGluLeuPheValGln 51
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
202 GCATCTCTGTGATCTGTTTGAA 222
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
52 SerGluIlePheProLeuGlu 58

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-914-853-11

seq_documentation_block:
; Sequence 11, Application US/08914853
; Patent No. 598186
; GENERAL INFORMATION:
; APPLICANT: Lou, Lillian Lien-Li
; APPLICANT: Barnett, Jimmy Wayne
; TITLE OF INVENTION: Cloning and Expression of Human GMP
; TITLE OF INVENTION: Synthetase, its use in Screening for
; TITLE OF INVENTION: Human
; TITLE OF INVENTION: GMP Synthetase and Inhibitors of Human
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Syntex (USA) Inc.
; STREET: 3401 Hillview Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA

```

```

1 ZIP: 94304
2 COMPUTER READABLE FORM:
3 MEDIUM TYPE: Floppy disk
4 COMPUTER: IBM PC compatible
5 OPERATING SYSTEM: PC-DOS/MS-DOS
6 SOFTWARE: PatentIn Release #1.0, Version #1.25
7 CURRENT APPLICATION DATA:
8 APPLICATION NUMBER: US/08/914,853
9 FILING DATE:
10 CLASSIFICATION:
11 PRIOR APPLICATION DATA:
12 APPLICATION NUMBER: 08/461,489
13 FILING DATE:
14 ATTORNEY/AGENT INFORMATION:
15 NAME: Peries, Rohan
16 REGISTRATION NUMBER: 35,752
17 REFERENCE/DOCKET NUMBER: 28060
18 TELECOMMUNICATION INFORMATION:
19 TELEPHONE: (415)-852-1698
20 TELEFAX: (415)-496-3529
21 INFORMATION FOR SEQ ID NO: 11:
22 SEQUENCE CHARACTERISTICS:
23 LENGTH: 693 amino acids
24 TYPE: amino acid
25 STRANDEDNESS: single
26 TOPOLOGY: linear
27 MOLECULE TYPE: protein
28 HYPOTHEetical: NO
29 US-08-914-853-11...
30
31 alignment_scores:
32 ... Quality: 57.50 Length: 57
33 ... Ratio: 1.643 Gaps: 1
34 Percent Similarity: 61.404 Percent Identity: 24.561
35
36 alignment_block:
37 US-09-049-696-1 x US-08-914-853-11 ..
38 Align: seq 1/1 to: US-08-914-853-11 from: 1 to: 693
39
40 79'GCCCTGAGTAATCTACTCATTCAGTCGACACAAATGC..... 117
41 ||||| ||||| : |||||: ||||| |||
42 2 AlaLeuCyAsnGlyAspSerLysLeuGluAsnAlaGlyGlyAspLeuLy 18
43
44 118 .....TATGAAGGCATTGTCGTTCCATTCGACCCCAATG 151
45 ||||| ||||| ||||| ||||| : ||||| : |||||
46 18'aspGlyHisHisTyrGluGlyAlaValIleLeuAspAlaGlyA 35
47
48 152 TGCCAGAGATGAACACTCATTCACAAATAAAGGACATGGTGACCCG 201
49
50 35'laGlnTyGlyLysValIleaspArgValargValGluLeuPheValGln 51
51
52 202 GCATCTCTGTATCTCTTTGAA 222
53 .....: |||||
54 52 SerGluIlePheProLeuGlu 58
55
56 seq_name: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US95-03934A-11
57
58 seq_documentation_block:
59 : Sequence 11, Application PC/TUS9503934A
60 : GENERAL INFORMATION:
61 : APPLICANT: Syntex (USA) Inc.
62 : TITLE OF INVENTION: Cloning and Expression of Human GMP
63 : TITLE OF INVENTION: Synthetase, its use in Screening for In
64 : TITLE OF INVENTION: GMP Synthetase and Inhibitors of Human
65 : NUMBER OF SEQUENCES: 11
66 : COMPUTER READABLE FORM:
67 : MEDIUM TYPE: Floppy disk
68 : COMPUTER: IBM PC compatible
69 : OPERATING SYSTEM: PC-DOS/MS-DOS
70 : SOFTWARE: PatentIn Release #1.0, Version #1.25
71 : CURRENT APPLICATION DATA:
72
73

```


608 rSerGlnProCysArgLeuArgGlyThrCysGlnAspProAspAsnAlaT 625
122 TCATACCCATCTCTGTTCACTGAATGAGTGAATTAATCAGGCGCCCTTC 73
625 yrLeuCysPheCysLeuLys.....GlyThrThrGlyProAsn 637
72 TAGAAGGTGAAGAATCAAGATGAACACAGACTCTTAATGCCCCCATG 23
638 CysGluIleAsnLeuAspAsp.....Cy 645
22 CTGTACATCTCCCTGTGAT 4
645 salaserSerProCysAsp 651

seq_name: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.us-08-532-384-20

seq_documentation_block:

; Sequence 20, Application US/08532384
; Patent No. 6083904

; GENERAL INFORMATION:

; APPLICANT: Artavanis-Tsakonas, S. et al.

; TITLE OF INVENTION: Therapeutic And Diagnostic Methods

; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And

; TITLE OF INVENTION: Nucleic Acids

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/532,384

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/083,590

; FILING DATE: 25-JUN-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 7326-015

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212 790-9090

; TELEFAX: 212 8698864/9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2556 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: unknown

; MOLECULE TYPE: peptide

US-08-532-384-20

alignment_scores:

Quality: 57.00 Length: 73

Ratio: 1.676 Gaps: 4

Percent Similarity: 46.575 Percent Identity: 26.027

alignment_block:

US-09-049-696-1/rev x US-08-532-384-20 ..

Align seg 1/1 to: US-08-532-384-20 from: 1 to: 2556

198 GGTACCATCTCCTTTATTTGTTGAATGAGTGTTCATCTTGTGCACAT 149

||||||| :|||
597 GlyHisHisCysGluThrAsnIleAsnGluCys.....Se 608
148 TGGGTGCGATGCAAC.....GACAATGCCT 123
608 rSerGlnProCysArgLeuArgGlyThrCysGlnAspProAspAsnAlaT 625
122 TCATAGCCATTGTTTTCAGCTGAATGAGTGAATTAATCAGGCGCCCTTC 73
625 yrLeuCysPheCysLeuLys.....GlyThrThrGlyProAsn 637
72 TAGAAGGTGAAGAATCAAGATGAACACAGAACTCTTAATGCCCCCATG 23
638 CysGluIleAsnLeuAspAsp.....Cy 645
22 CTGTACATCTCCCTGTGAT 4
645 salaserSerProCysAsp 651

DR WPI: 2001-441847/47.
XX Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer -
XX
XX
PS Claim 2; Page 440-443; 472pp; English.
XX
CC The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be
CC used to treat disorders associated with decreased expression by
CC rectifying mutations or deletions in a patient's genome that affect the
CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host
CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.
XX
SQ Sequence 914 AA;

alignment_scores:
Quality: 328.00 Length: 66
Ratio: 4.970 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-1 x AAM24514 ..

Align seg 1/1 to: AAM24514 from: 1 to: 914

25 ATGGGGCCATTAAAGATTCCTGCTTCATCTTGATTCACCTCTCTAGA 74
|||||
1 MetGlyProPheLysSerSerValPheLeuLeuLeuHisLeuLeuG1 17
75 AGGGCCCTGAGTAATTCACATTCACCTTGACCAACAATGCTATGAAG 124
|||||
17 uGlyAlaLeuSerAsnSerLeuLeuGlnLeuAsnAsnGlyTyrGluG 34

125 GCATTGCTGCTTGCATTCACCCCAATGTGCCAGAGATCAACACTCATTT 174
|||||
34 lylleValValAlaIleAspProAsnValProGluAspGlnThrLeuLeu 50
175 CAACAAATAAAGGACATGCTGACCCAGGCATCTCTGTATCTGTTTGA 222
|||||
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGlu 66

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB73716

seq_documentation_block:

ID AAB73716 standard; Protein: 914 AA.

XX AAB73716;

XX AC

XX 11-SEP-2001 (first entry)

XX Human CLC1 protein, SEQ ID NO:2.

XX Human CLC1; goblet cell; mouse Gob-5 orthologue; drug screening;

KW Expression inhibition; antisense therapy; gene therapy;
KW bronchial obstructive pulmonary disease; bronchial asthma; antiasthmatic.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Misc-difference 152 /note= "Encoded by AGC in AAH46124"
XX W230138530-AA.
XX 31-MAY-2001.
XX
XX 22-NOV-2000; 2000WO-JP08232.
XX 24-NOV-1999; 99JP-0333479.
XX 27-APR-2000; 2000JP-0127589.
XX (TAKE) TAKEDA CHEM IND LTD.
XX Akanishi A, Morita S;
XX WPI: 2001-355935/37.
XX N-PSDB; AAH46102, AAH46124.
XX
XX New antisense nucleotide, useful for treatment and prevention of
XX bronchial asthma and chronic obstructive pulmonary disease -
XX Claim 2; Page 76-80; 104pp; Japanese.
XX
XX The invention relates to an antisense nucleotide targetted to the mouse
XX Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
XX the CLC1 gene (coding sequence shown in AAH46102). The invention also
XX relates to an antibody specific for the Gob-5 protein, medical and
XX diagnostic compositions containing the antisense nucleotide or the
XX antibody, and methods and kits for screening for compounds which inhibit
XX the protein. Gob-5 and CLC1 are proteins expressed by goblet cells.
XX The antisense oligonucleotides and antibody are therefore useful for the
XX treatment and prevention of bronchial asthma and chronic obstructive
XX pulmonary disease. The present sequence represents human CLC1 protein.
XX
SQ Sequence 914 AA;

alignment_scores:

Quality: 328.00 Length: 66
Ratio: 4.970 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-1 x AAB73716 ..

Align seg 1/1 to: AAB73716 from: 1 to: 914

25 ATGGGGCCATTAAAGATTCCTGCTTCATCTTGATTCACCTCTCTAGA 74
|||||
1 MetGlyProPheLysSerSerValPheLeuLeuLeuHisLeuLeuG1 17
75 AGGGCCCTGAGTAATTCACATTCACCTTGACCAACAATGCTATGAAG 124
|||||
17 uGlyAlaLeuSerAsnSerLeuLeuGlnLeuAsnAsnGlyTyrGluG 34
125 GCATTGCTGCTTGCATTCACCCCAATGTGCCAGAGATCAACACTCATTT 174
|||||
34 lylleValValAlaIleAspProAsnValProGluAspGlnThrLeuLeu 50
175 CAACAAATAAAGGACATGCTGACCCAGGCATCTCTGTATCTGTTTGA 222
|||||
51 GlnGlnIleLysAspMetValThrGlnAlaSerLeuTyrLeuPheGlu 66

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB74733

seq_documentation_block:

ID XX AAB74733 standard; Protein; 914 AA.

AC XX AAB74733;

CC XX 12-JUN-2001 (first entry)

DE XX Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.

KW Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;

KW dermatological; immunosuppressive; anti-inflammatory; anti-HIV;

KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;

KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;

KW antitumor; antiparkinsonian; antimicrobial; vulnary; gene therapy;

KW immune disorder; hyperproliferative disorder; cardiovascular disease;

KW cancer; angiogenic disorder; neurological disorder; infectious disease;

KW wound healing; regeneration; chemotaxis; chromosome 1.

OS Homo sapiens.

XX WO200112775-A2.

PN 22-FEB-2001.

XX 16-AUG-2000; 2000WO-US22325.

XX 17-AUG-1999; 99US-0149182.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, NI J, Florence KA, Fiscella M, Wei P, Baker KP;

PI Birse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;

XX WPI; 2001-147550/15.

DR N-PSDB; AAF81787.

XX Nucleic acids encoding 25 human secreted polypeptides, useful for

PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's

PT disease and diabetic retinopathy -

XX Claim 11; Page 459-460; 485pp; English.

XX AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733

CC to AAB74772. Human secreted proteins can have activities based on the

CC tissues and cells they are expressed in. Example of activities include:

CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;

CC anti-inflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;

CC vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic;

CC anticonvulsant; antitumor; antiparkinsonian; antimicrobial; and

CC vulnary. Human secreted proteins can be used in gene therapy and

CC vaccine. Human secreted protein nucleotide sequences (NAM1) and proteins

CC (PEP1) may be used in the prevention, diagnosis and treatment of diseases

CC associated with inappropriate polypeptide expression. For example, NAM1

CC and pep1 may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patients genome

CC that affect the activity of proteins by expressing inactive proteins or

CC to supplement the patients own production of polypeptides. Disorders that

CC may be prevented, diagnosed and/or treated include immune disorders,

CC hyperproliferative disorders (e.g. cancers), cardiovascular diseases,

CC angiogenic disorders, neurological disorders, infectious diseases and/or

CC for promoting wound healing; regeneration and /or chemotaxis. AAF81778 to

CC AAF81786 and AAB74732 represent sequences used in the exemplification of

CC the present invention.

XX SQ Sequence 914 AA;

alignment_scores:

Quality: 328.00 Length: 66

Ratio: 4.970 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

us-09-049-696-1 x AAB74733

Allig seg 1/1 to: AAB74733 From: 1 to: 914

25 ATCGGGCCCAATTAAGAGTTCCTGCTTCATCTTCACCTTCCTAGTA 74

1 MetGlyProPheLeuSerValPheLeuLeuLeuLeuLeuLeuLeuLeu 17

75 AGGGGGCCCTGAGCTAATTCATTCAGCTGAGCAACAATGGCTATGAAG 124

7 uGlyAlaLeuSerAsnSerLeuLeuLeuLeuLeuLeuLeuLeuLeu 34

25 GCATTTCGTTGCAATCGACCCCAATGTGCCAGACATGAACACTCAT 174

34 LylleValValAlaLeAspProAsnValProGluAspGluThrLeuLeu 50

175 CACAATAAAGGACATGFGACCCAGGCATCTCTATCTATCTGTTTCAA 222

1 GlnGlnLeuLeuAspMetValThrGlnAlaSerLeuThrLeuPheGlu 66

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74822

seq_documentation_block:

ID AAB74822 standard; Protein; 913 AA.

XX AAB74822;

XX 13-JUN-2001 (first entry)

XX Marine ICACC-1 protein sequence.

XX ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;

KW Interleukin 9 induced calcium activated chloride channel; IL-9;

KW calcium activated chloride channel; anti-allergic; anti-asthmatic;

KW anti-inflammatory; immunomodulatory; cystic fibrosis;

KW inflammatory bowel disease; autoimmune disease.

XX Mus sp.

OS WO9944620-A1.

PN 10-SEP-1999.

XX 03-MAR-1999; 99WO-US04703.

XX 03-MAR-1998; 98US-0076815.

XX (MACA-) MACAININ PHARM INC.

XX Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;

PI Nicolaides NC, Zhou Y, Dong Q;

XX WPI; 1999-550979/46.

DR N-PSDB; AAF81925.

XX New nucleic acid encoding calcium activated chloride channel, used to

PT identify, e.g. specific modulators for treating atopic allergy -

XX Claim 12; Fig 2; 75pp; English.

XX The present sequence represents the murine interleukin 9 (IL-9) induced

CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins

CC have anti-allergic, anti-asthmatic, anti-inflammatory and

CC immunomodulatory activities. Compounds (A) that downregulate ICACC are

CC used to alleviate asthma (or more generally atopic allergy), while those

CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and

CC inflammatory bowel disease (IBD) (or other autoimmune diseases).

CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma

CC (increased levels) or IBD (reduced levels), also for monitoring

CC treatment of these conditions. The ICACC proteins can be used:

CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay

CC reagents and (b) as therapeutic (A); (ii) as fragments or mutants, as

CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to

CC identify modulators and binding partners. ICACC polynucleotides can be

CC used to generate transgenic animals or recombinant cells, used to screen
 CC for antagonists, also as a source of therapeutic antisense agents or
 CC diagnostic probes (for quantifying mRNA expression, e.g. for
 CC identification of modulators).

XX Sequence 913 AA;

alignment_scores:
 Quality: 257.00 Length: 66
 Ratio: 4.213 Gaps: 0
 Percent Similarity: 92.424 Percent Identity: 80.303

alignment_block:

US-09-049-696-1 x AAB74822

Align seg 1/1 to: AAB74822 from: 1 to: 913

```

25 ATGGGGCCCATTTAAGAGTCTCTGTTTCATCTTGATTCCTACCTTCTAGA 74
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MetGluSerLeuLysSerProValPheLeuLeuLeuLeuLeuLeuG1 17

75 AGGGGCCCTGAGTAATTCACCTCATTCAGCTGAACAAATGGCTATGAAG 124
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17 uGlyValLeuSerGluSerLeuLeuLeuGlnLeuAsnAsnGlyTyrGluG 34

125 GCATTGCTGCTGCATTCGACCCCAATGTGCCAGAGATGAACACTCATTT 174
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 lylleValleAlaIleAspMetValThrGlnAlaSerProTyrLeuPheGlu 50

175 CAACAAATAAAGGACATCGTGACCCAGGCATCTCTGATCTCTTTGAA 222
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 GlnHisIleLysAspMetValThrGlnAlaSerProTyrLeuPheGlu 66

```

seq_name: /SDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB73715

seq_documentation_block:

ID AAB73715 standard; Protein: 913 AA.

XX AAB73715;

AC 11-SEP-2001 (first entry)

XX Mouse Gob-5 protein, SEQ ID NO:1.

XX Mouse Gob-5; murine; goblet cell; human CLCA1 orthologue; drug screening;
 KW expression inhibition; antisense therapy; gene therapy; bronchial asthma;
 KW chronic obstructive pulmonary disease; antiasthmatic.

XX Mus sp.

XX W0200138530-A1.

XX 31-MAY-2001.

XX 22-NOV-2000; 2000WO-JP08232.

XX 24-NOV-1999; 99JP-0333479.

XX 27-APR-2000; 2000JP-0127589.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Nakanishi A, Morita S;

XX WPI; 2001-355935/37.

XX N-PSDB; AAH46101, AAH46120.

XX New antisense nucleotide, useful for treatment and prevention of
 PT bronchial asthma and chronic obstructive pulmonary disease -

XX Claim 1; Page 72-76; 104pp; Japanese.

XX The invention relates to an antisense nucleotide targetted to the mouse

CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
 CC the CLCA1 gene (coding sequence shown in AAH46102). The invention also
 CC relates to an antibody specific for the Gob-5 protein, medical and
 CC diagnostic compositions containing the antisense nucleotide or the
 CC antibody, and methods and kits for screening for compounds which inhibit
 CC the protein, Gob-5 and CLCA1 are proteins expressed by goblet cells.
 CC the antisense oligonucleotides and antibody are therefore useful for the
 CC treatment and prevention of bronchial asthma and chronic obstructive
 CC pulmonary disease. The present sequence represents mouse Gob-5 protein.

XX Sequence 913 AA;

alignment_scores:

Quality: 257.00 Length: 66
 Ratio: 4.213 Gaps: 0
 Percent Similarity: 92.424 Percent Identity: 80.303

alignment_block:

US-09-049-696-1 x AAB73715

Align seg 1/1 to: AAB73715 from: 1 to: 913

```

25 ATGGGGCCCATTTAAGAGTCTCTGTTTCATCTTGATTCCTACCTTCTAGA 74
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1 MetGluSerLeuLysSerProValPheLeuLeuLeuLeuLeuG1 17

75 AGGGGCCCTGAGTAATTCACCTCATTCAGCTGAACAAATGGCTATGAAG 124
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
17 uGlyValLeuSerGluSerLeuLeuGlnLeuAsnAsnGlyTyrGluG 34

125 GCATTGCTGCTGCATTCGACCCCAATGTGCCAGAGATGAACACTCATTT 174
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 lylleValleAlaIleAspMetValThrGlnAlaSerProTyrLeuPheGlu 50

175 CAACAAATAAAGGACATCGTGACCCAGGCATCTCTGATCTCTTTGAA 222
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
51 GlnHisIleLysAspMetValThrGlnAlaSerProTyrLeuPheGlu 66

```

seq_name: /SDS2/gcgdata/geneseq/geneseq/AA2000.DAT: AAY66749

seq_documentation_block:

ID AAY66749 standard; protein: 919 AA.

XX AAY66749;

XX 05-APR-2000 (first entry)

XX Membrane-bound protein PRO1124.

XX Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
 KW pharmaceutical; receptor immunoadhesin; gene mapping.

XX Homo sapiens.

XX AA9963088-A2.

XX 01-DEC-1999.

XX 02-JUN-1999; 99WO-US12252.

XX 02-JUN-1998; 98US-0087607.

XX 02-JUN-1998; 98US-0087609.

XX 02-JUN-1998; 98US-0087759.

XX 03-JUN-1998; 98US-0087827.

XX 03-JUN-1998; 98US-0088021.

XX 03-JUN-1998; 98US-0088025.

XX 04-JUN-1998; 98US-0088028.

XX 04-JUN-1998; 98US-0088029.

XX 04-JUN-1998; 98US-0088030.

XX 04-JUN-1998; 98US-0088033.

XX 04-JUN-1998; 98US-0088326.

XX 05-JUN-1998; 98US-0088167.

PR 05-JUN-1998; 98US-0088202.
 PR 05-JUN-1998; 98US-0088212.
 PR 05-JUN-1998; 98US-0088217.
 PR 09-JUN-1998; 98US-0088655.
 PR 10-JUN-1998; 98US-0088722.
 PR 10-JUN-1998; 98US-0088730.
 PR 10-JUN-1998; 98US-0088734.
 PR 10-JUN-1998; 98US-0088738.
 PR 10-JUN-1998; 98US-0088740.
 PR 10-JUN-1998; 98US-0088741.
 PR 10-JUN-1998; 98US-0088742.
 PR 10-JUN-1998; 98US-0088810.
 PR 10-JUN-1998; 98US-0088824.
 PR 10-JUN-1998; 98US-0088825.
 PR 10-JUN-1998; 98US-0088826.
 PR 11-JUN-1998; 98US-0088858.
 PR 11-JUN-1998; 98US-0088861.
 PR 11-JUN-1998; 98US-0088863.
 PR 11-JUN-1998; 98US-0088876.
 PR 12-JUN-1998; 98US-0089090.
 PR 12-JUN-1998; 98US-0089105.
 PR 16-JUN-1998; 98US-0089440.
 PR 16-JUN-1998; 98US-0089512.
 PR 16-JUN-1998; 98US-0089514.
 PR 17-JUN-1998; 98US-0089532.
 PR 17-JUN-1998; 98US-0089538.
 PR 17-JUN-1998; 98US-0089598.
 PR 17-JUN-1998; 98US-0089599.
 PR 17-JUN-1998; 98US-0089600.
 PR 17-JUN-1998; 98US-0089653.
 PR 18-JUN-1998; 98US-0089801.
 PR 18-JUN-1998; 98US-0089907.
 PR 18-JUN-1998; 98US-0089908.
 PR 19-JUN-1998; 98US-0089947.
 PR 19-JUN-1998; 98US-0089948.
 PR 19-JUN-1998; 98US-0089952.
 PR 22-JUN-1998; 98US-0090246.
 PR 22-JUN-1998; 98US-0090252.
 PR 22-JUN-1998; 98US-0090254.
 PR 23-JUN-1998; 98US-0090349.
 PR 23-JUN-1998; 98US-0090355.
 PR 24-JUN-1998; 98US-0090429.
 PR 24-JUN-1998; 98US-0090431.
 PR 24-JUN-1998; 98US-0090435.
 PR 24-JUN-1998; 98US-0090444.
 PR 24-JUN-1998; 98US-0090445.
 PR 24-JUN-1998; 98US-0090461.
 PR 24-JUN-1998; 98US-0090472.
 PR 24-JUN-1998; 98US-0090535.
 PR 24-JUN-1998; 98US-0090538.
 PR 24-JUN-1998; 98US-0090540.
 PR 24-JUN-1998; 98US-0090557.
 PR 25-JUN-1998; 98US-0090676.
 PR 25-JUN-1998; 98US-0090678.
 PR 25-JUN-1998; 98US-0090688.
 PR 25-JUN-1998; 98US-0090690.
 PR 25-JUN-1998; 98US-0090691.
 PR 25-JUN-1998; 98US-0090694.
 PR 25-JUN-1998; 98US-0090695.
 PR 25-JUN-1998; 98US-0090696.
 PR 25-JUN-1998; 98US-0090696.
 PR 26-JUN-1998; 98US-0090862.
 PR 26-JUN-1998; 98US-0090863.
 PR 01-JUL-1998; 98US-0091358.
 PR 01-JUL-1998; 98US-0091360.
 PR 01-JUL-1998; 98US-0091544.
 PR 02-JUL-1998; 98US-0091478.
 PR 02-JUL-1998; 98US-0091486.
 PR 02-JUL-1998; 98US-0091519.
 PR 02-JUL-1998; 98US-0091626.
 PR 02-JUL-1998; 98US-0091628.
 PR 02-JUL-1998; 98US-0091633.
 PR 02-JUL-1998; 98US-0091646.

PR 02-JUL-1998; 98US-0091673.
 PR 07-JUL-1998; 98US-0091978.
 PR 07-JUL-1998; 98US-0091982.
 PR 09-JUL-1998; 98US-0092182.
 PR 10-JUL-1998; 98US-0092472.
 PR 20-JUL-1998; 98US-0093339.
 PR 30-JUL-1998; 98US-0094651.
 PR 04-AUG-1998; 98US-0095282.
 PR 04-AUG-1998; 98US-0095285.
 PR 04-AUG-1998; 98US-0095301.
 PR 04-AUG-1998; 98US-0095302.
 PR 04-AUG-1998; 98US-0095318.
 PR 04-AUG-1998; 98US-0095321.
 PR 04-AUG-1998; 98US-0095325.
 PR 10-AUG-1998; 98US-0095916.
 PR 10-AUG-1998; 98US-0095929.
 PR 10-AUG-1998; 98US-0096012.
 PR 11-AUG-1998; 98US-0096143.
 PR 11-AUG-1998; 98US-0096146.
 PR 12-AUG-1998; 98US-0096329.
 PR 17-AUG-1998; 98US-0096757.
 PR 17-AUG-1998; 98US-0096766.
 PR 17-AUG-1998; 98US-0096768.
 PR 17-AUG-1998; 98US-0096773.
 PR 17-AUG-1998; 98US-0096791.
 PR 17-AUG-1998; 98US-0096867.
 PR 17-AUG-1998; 98US-0096891.
 PR 17-AUG-1998; 98US-0096894.
 PR 17-AUG-1998; 98US-0096895.
 PR 17-AUG-1998; 98US-0096897.
 PR 18-AUG-1998; 98US-0096949.
 PR 18-AUG-1998; 98US-0096950.
 PR 18-AUG-1998; 98US-0096959.
 PR 18-AUG-1998; 98US-0096960.
 PR 18-AUG-1998; 98US-0097022.
 PR 19-AUG-1998; 98US-0097141.
 PR 20-AUG-1998; 98US-0097218.
 PR 24-AUG-1998; 98US-0097661.
 PR 26-AUG-1998; 98US-0097951.
 PR 26-AUG-1998; 98US-0097952.
 PR 26-AUG-1998; 98US-0097954.
 PR 26-AUG-1998; 98US-0097955.
 PR 26-AUG-1998; 98US-0097971.
 PR 26-AUG-1998; 98US-0097974.
 PR 26-AUG-1998; 98US-0097978.
 PR 26-AUG-1998; 98US-0097979.
 PR 26-AUG-1998; 98US-0097986.
 PR 26-AUG-1998; 98US-0098014.
 PR 31-AUG-1998; 98US-0098525.
 PR 16-SEP-1998; 98US-0100634.
 PR 12-JAN-1999; 98US-0115565.

(GETH) GENENTECH INC.

Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
 Wood WL, Yuan J;

WPI; 2000-072883/06.
 N-PSDB; AA25095.

Membrane-bound proteins and related nucleotide sequences

claim 12; Fig.274; 822pp; English.

The invention provides membrane-bound PRO polypeptides and polynucleotides encoding them. The PRO sequences of the invention were identified based on extracellular domain homology screening. The PRO sequences have homology with proteins including LDL receptors, TIE ligands and various enzymes. The membrane-bound proteins and receptor molecules are useful as pharmaceutical and diagnostic agents. Receptor molecules, for instance, can be used as therapeutic agents to block receptor-ligand interactions. The membrane-bound proteins can also be employed for screening of potential peptide or small molecule inhibitors


```
XX OS Homo sapiens.
XX PN WO2000061612-A2.
XX PD 19-OCT-2000.
XX PF 03-APR-2000; 2000WO-US08896.
XX PR 02-APR-1999; 99US-0285479.
XX PR 17-DEC-1999; 99US-0466396.
XX PR 30-DEC-1999; 99US-0476496.
XX PR 10-JAN-2000; 2000US-0480884.
XX PR 22-FEB-2000; 2000US-0510376.
XX PA (CORI-) CORIXA CORP.
XX PJ Wang T, Fan L;
XX DR WPI; 2000-628399/60.
XX DR N-PSDB; AAC65897.
XX PT Isolated polypeptide comprising an immunogenic portion of a lung tumor
XX PT protein is used for detecting and monitoring progression of lung cancer
XX PT in a patient.
XX PS Claim 3; Page 178-180; 261pp; English.
XX CC This invention describes a novel isolated polypeptide (I) which
XX CC comprising an immunogenic portion of a lung tumor protein or variant (P2)
XX CC which have cytostatic activity. The polypeptides and polynucleotides are
XX CC used in compositions and vaccines to inhibit the development of cancer,
XX CC especially lung cancer, in a patient. Methods described in the invention
XX CC can be used to monitor the progression of a cancer by carrying out the
XX CC detection at subsequent time points and comparing the results from the
XX CC different time points. CD4+ and/or CD8+ T-Cells isolated from a patient
XX CC are treated with P2, polynucleotides encoding P2 or antigen presenting
XX CC cells expressing P2 and then administered to the patient to inhibit
XX CC development of cancer.
XX SQ Sequence 791 AA;

alignment_scores:
  Quality: 148.50      Length: 78
  Ratio: 2.517        Gaps: 3
  Percent Similarity: 75.641  Percent Identity: 43.590

alignment_block:
US-09-049-696-1 x AAB11325 ..

Align seg 1/1 to: AAB11325 from: 1 to: 791

4 ATCACAGGAGATGTACAGCAATGGGCGCATTTAAGAGTTCTGTGTCAT 53
   : : : : : : : : : : : : : : : : : : : : : : : : : :
1 MethrGlnArgSerIleAla...GlyProlleCysAsnLeuLysPheVa 16
   : : : : : : : : : : : : : : : : : : : : : : : : : :
54 CTTGATTCCTTCACCTCTAGAAGGGGCCCTGAGTAATTCATC 96
   : : : : : : : : : : : : : : : : : : : : : : : : : :
16 lThrLeuLeu.....ValAlaLeuSerSerGluLeuProPheL 29
   : : : : : : : : : : : : : : : : : : : : : : : : : :
97 .....ATTACGCTGAACACATGGCTATGAGGCATTCGTT 135
   : : : : : : : : : : : : : : : : : : : : : : : : : :
29 euGlyAlaGlyValGlnLeuGlnAspGlyTyrAsnGlyLeuLeuIle 45
   : : : : : : : : : : : : : : : : : : : : : : : : : :
136 GCAATCGACCCCAATGTCGAGAGATGAACACATTCATCAACAAATAA 185
   : : : : : : : : : : : : : : : : : : : : : : : : : :
46 AlaIleAsnProGlnValProGluAsnGlnAsnLeuIleSerAsnIlely 62
   : : : : : : : : : : : : : : : : : : : : : : : : : :
186 GGACATGGTGACCCAGGCATCTCTGTATCTGTTT 219
   : : : : : : : : : : : : : : : : : : : : : : : : : :
62 sGluMetIleThrGluAlaSerPheTyrLeuPhe 73
```

This Page Blank (uspto)

OM of: US-09-049-696-15 to: Issued Patents_AA.* out_format : pfs
Date: Mar 30, 2002 2:33 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+g2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09049696/runat_28032002_145238_2012/app_query.fasta_1.12579
-DB=Issued Patents_AA -QFMT=fastan -SUFFIX=ra1 -GAPEXT=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -XGAPEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcet
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1_109 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-15
Query length: 154
Database: Issued Patents_AA.*
Database sequences: 212252
Database length: 22503292
Search time (sec): 362.110000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-469-667-9	+	240.00	581.38	27e-25	228
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.ppt:US-08-469-667-9	+	240.00	581.38	2.7e-25	228
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-08-808-599A-24	+	63.50	127.15	1.08	1160
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-08-415-655-5	+	60.00	129.95	2.51	348
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-08-415-655-13	+	60.00	129.95	2.51	348
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-08-415-655-15	+	60.00	129.95	2.51	348
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-080-897-2	+	59.50	116.51	3.92	1248
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-323-735-2	+	59.50	116.51	3.92	1248
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-08-899-593-3	+	59.50	116.01	3.96	1315
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-103-429A-3	+	59.00	119.68	4.14	786
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-103-429A-4	+	59.00	119.45	4.16	805
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-252-966B-12	+	58.50	113.99	5.39	1253
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-252-966B-18	+	58.50	113.93	5.40	1261
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-08-922-635-22	+	58.00	114.25	6.11	1070
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-378-255-2	+	57.50	121.34	5.89	447
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-715-336-2	+	57.50	121.34	5.89	447
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-251-372-2	+	57.50	120.76	5.97	475
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-552-351-2	+	57.50	118.95	6.23	574
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-08-899-437-23	+	57.50	117.11	6.50	696
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-126-121-23	+	57.50	117.11	6.50	696
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-08-899-437-6	+	57.50	116.79	6.55	720
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-126-121-6	+	57.50	116.79	6.55	720
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-141-212-8	+	57.50	114.90	6.85	878
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-561-138-8	+	57.50	114.90	6.85	878
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-141-212-6	+	57.50	114.89	6.85	879
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-561-138-6	+	57.50	114.89	6.85	879
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-916-917-12	+	57.50	114.87	6.85	880
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-225-170-12	+	57.50	114.87	6.85	880
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-378-255-6	+	57.50	114.87	6.85	880
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-141-212-2	+	57.50	114.87	6.85	880
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-141-212-4	+	57.50	114.87	6.85	880
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-552-351-4	+	57.50	114.87	6.85	880
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-251-372-6	+	57.50	114.87	6.85	880
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-561-138-2	+	57.50	114.87	6.85	880
/cgn2_6/ptodata/2/iaa/6A_COMB.ppt:US-09-561-138-4	+	57.50	114.87	6.85	880
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-178-477B-32	+	57.00	118.49	7.17	529
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-09-715-336-6	+	56.00	117.18	9.60	468
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-09-537-357-53	+	56.00	110.59	11.20	933
/cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-313-200-1	+	56.00	110.59	11.20	933
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.ppt:US-08-469-667-9	+	56.00	110.59	11.20	933

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.ppt:US-08-469-667-9
seq_documentation_block:
Sequence 9, Application US/08469667
Patent No. 5733748
GENERAL INFORMATION:
APPLICANT: Yu, Guo-Liang
APPLICANT: Rosen, Craig
TITLE OF INVENTION: Colon Specific Genes and Proteins
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
ADDRESSEE: Stewart & Olstein
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: NJ
COUNTRY: USA
ZIP: 07068-1739
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,667
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Ferraro, Gregory D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-435
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 228 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-469-667-9
alignment_scores:
Quality: 240.00 Length: 47
Ratio: 5.106 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-049-696-15 x US-08-469-667-9
Align seg 1/1 to: US-08-469-667-9 from: 1 to: 228
1 ACAGATCTTTTCATTGCTATTACGCTCTTGTATAGGTCGATCTCAATC 50
182 ThrAspLeuPheAlaIleAlaValAspLysLeuLys 198
51 AGAATATCCAACTTGCACGAGTATCTTTGTTTATCTCCACAGACTC 100
193 rGluLeuSerAsnIleAlaArgValSerLeuPheIleProGlnThr 215
101 CGCCAGAGACACCTAGTCTCTGATGAACGTCGTCTCTGTGT 141
215 roProGluThrProSerProAspGluThrSerAlaProCys 228
seq_name: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.ppt:US-08-469-667-9

```
seq_documentation_block:
; Sequence 9, Application PC/TUS9507289
; GENERAL INFORMATION:
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Rosen, Craig
; TITLE OF INVENTION: Colon Specific Genes and Proteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; ADDRESSEE: Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07289
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-265
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-07289-9

alignment_scores:
Quality: 240.00 Length: 47
Ratio: 5.106 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-15 x PCT-US95-07289-9
Align seg 1/1 to: PCT-US95-07289-9 from: 1 to: 228

1 ACAGATCTTTTCATTGCTATTTCAGGCTGTGATAAGTCGATCTGAATC 50
182 ThrAspLeuPheIleAlaIleGlnAlaValAspLysValAspLeuLys 198
51 AGAATATCAACATTCACGAGTATCTTTGTTATTCTCCACAGACTC 100
198 rGluIleSerAsnIleAlaIleArgValSerLeuPheIleProGlnThrP 215
101 CGCCAGAGACACCTAGCTGATGAACGCTGCTCTCTGT 141
215 roProGluThrProSerProAspGluThrSerAlaProCys 228

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-808-599A-24

seq_documentation_block:
; Sequence 24, Application US/080808599A
; Patent No. 6111089
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin, Trophinin-Assisting
; TITLE OF INVENTION: Proteins and Methods to Inhibit Implantation
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/808,599A
; FILING DATE: 28-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/317,522
; FILING DATE: 04-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/439,818
; FILING DATE: 12-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 2256
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1160 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-808-599A-24

alignment_scores:
Quality: 63.50 Length: 41
Ratio: 2.352 Gaps: 1
Percent Similarity: 65.854 Percent Identity: 39.024

alignment_block:
US-09-049-696-15/rev x US-08-808-599A-24
Align seg 1/1 to: US-08-808-599A-24 from: 1 to: 1160

146 TTAGGACAAGAGCAGACGCTTTTCATCAGGACTAGGTGCTCTGCGCGAGT 97
888 LeuAsnHisAsnAlaAspPheAsnGlyLeuGlyLeuGlyAsnSerAlaGlyPh 904
96 CTGTGGAGGAATAACAAGATACCTCGTCAATGTTGGATATTCTGATT 47
904 eAsnGlyGlyLeuAsnThrAsnThr.....AspP 914
46 TCAGATCGACCTTATCAACAGCC 24
914 heGlyGlyGluLeuGlyThrSer 921

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-415-655-5

seq_documentation_block:
; Sequence 5, Application US/08415655
; Patent No. 6025480
; GENERAL INFORMATION:
; APPLICANT: Massague, Joan
; APPLICANT: Lee, Mong-hong
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: P75KIP2, A CYCLIN-DEPENDENT KINASE INHIBITOR AND USES OF
; TITLE OF INVENTION: SAME
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
```



```

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UW97-001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 343-4341
TELEFAX: (650) 343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1248 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-080-897-2

alignment_scores:
Quality: 59.50 Length: 54
Ratio: 2.125 Gaps: 1
Percent Similarity: 51.852 Percent Identity: 20.370

alignment_block:
US-09-049-696-15 x US-09-080-897-2 ..
Align seg 1/1 to: US-09-080-897-2 from: 1 to: 1248

40 GATGCTGAATTCAGAAATATCCCAACATATGCACGAGTATCTTTGTATTATCC 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
548 AspAlaLysLysLysLysMetAlaSerLeuSerAlaAlaAlaIleThrValPr 564
90 TCACAGACACTCG..... 102
|||||:|||||:
564 oProSerValProSerArgAlaProValProProAlaProProLeuProG 581
103 .....CCAGACACACTAGTCCTGATGAAACGTC 132
||| |||:|||||:
581 LysAspSerGlyThrIlelleProProProAlaProGlyAspSerThr 597
133 GTCCTGTGCTC 144
||||| |||
598 ThrProPro 601

seq_name: /cgn_2_6/ptodata/2/iaa/6B_COMB.pep:US-09-323-735-2

seq_documentation_block:
Sequence 2, Application US/09323735
Patent No. 6197932
GENERAL INFORMATION:
APPLICANT: King, Mary-Claire
APPLICANT: Lynch, Eric D.
APPLICANT: Lee, Ming
APPLICANT: Morrow, Jan E.
APPLICANT: Welch, Piri L.
APPLICANT: Leon, Pedro E.
TITLE OF INVENTION: Modulators of Actin
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 75 DENISE DRIVE
CITY: HILLSBOROUGH
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94010
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/323,735
FILING DATE:
CLASSIFICATION:

```

APPLICATION NUMBER: JP 8-242701
 FILING DATE: 26-APR-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 9-90170
 FILING DATE: 25-MAR-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Stephen A. Bent
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 049441/0112
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1315 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-899-595-3

alignment_scores:	
Quality:	59.50
Ratio:	2.125
Percent Similarity:	51.852
Percent Identity:	20.370
Length:	54
Gaps:	1

Align seg 1/1 to: US-09-323-735-2 from: 1 to: 1248

```

Align seg 1/1 to: US-08-899-595-3 from: 1 to: 1315
40 GATCTGAAATCAGAAATATCCAAATTCGACGAGTATCTTTGTTATTCC 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
60 AspAlaLysIysgluMetAlaSerLeuSerAlaAlaIleThrValPr 616
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 TCCACGAGACTCCG..... 102
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
616 oProSerValProSerArgAlaProValProProAlaProProLeuProG 633
   .....CCAGAGACACCTAGTCTCTGATGAAACGTCCT 132
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
633 LyspSerGlyThrIleProProProAlaProGlyAspSerThr 649
   .....GCTCCTTGTGCT 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
650 ThrProPro 653

```

```
seq_documentation_block:
; Sequence 3, Application US/08899595
; Patent No. 6111072
```

? GENERAL INFORMATION:
 ? APPLICANT: Narumiya, Shuh
 ? APPLICANT: Takahashi, No. 6111072uaki
 ? TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
 ? TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
 ? TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
 ? NUMBER OF SEQUENCES: 14
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Foley & Lardner
 ? STREET: 3000 K Street, N.W., Suite 500
 ? CITY: Washington
 ? STATE: D.C.
 ? COUNTRY: USA
 ? ZIP: 20007-5109
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Floppy disk
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: Patentin Release #1.0, Version #1.30
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/08/899,595
 ? FILING DATE: 24-JUL-1997
 ? CLASSIFICATION: 435
 ? PRIOR APPLICATION DATA:

seq_name: /cqn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-103-429A-3

```

1 seq_documentation_block:
2 Sequence 3, Application US/09103429A
3 Patent No. 6187558
4
5 GENERAL INFORMATION:
6
7 APPLICANT: Granados, Robert R
8
9 TITLE OF INVENTION: A NO. 6187558
10
11 TITLE OF INVENTION: CDNA and REL
12
13 NUMBER OF SEQUENCES: 4
14
15 CORRESPONDENCE ADDRESS:
16
17 ADDRESSEE: Brown, Pinnisi & Mi
18
19 STREET: 118 NO. 6187558th Tlog
20
21 CITY: Ithaca
22
23 STATE: NY
24
25 COUNTRY: USA
26
27 ZIP: 14850
28
29 COMPUTER READABLE FORM:
30
31 MEDIUM TYPE: Floppy disk
32
33 COMPUTER: IBM PC compatible
34
35 OPERATING SYSTEM: PC-DOS/MS-DOS
36
37 SOFTWARE: PatentIn Release #1.
38
39 CURRENT APPLICATION DATA:
40
41 APPLICATION NUMBER: US/09/103,

```

us-09-049-696-15.ra1

Tue Apr 2 09:39:45 2002

```

; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-2000
; TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 786 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; TISSUE TYPE: peritrophic membrane
; US-09-103-429A-3

alignment_scores:
    Quality: 59.00      Length: 28
    Ratio: 3.933       Gaps: 1
    Percent Similarity: 53.571      Percent Identity: 46.429

alignment_block:
US-09-049-696-15 x US-09-103-429A-3
..
Align seg 1/1 to: US-09-103-429A-3 from: 1 to: 786

67 GCACGAGTATCTTTGTTATTCCTCCACAGACTCCGCCAGACAGA..... 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
507 AlaAlaProSerThrValValProAlaThrProAlaThrAlaAl 523

112 .CCTAGTCTGTGAACGTCGCTCCTGTGCT 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
523 aProValProThrThrAlaIleProThrPro 534

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-103-429A-4

seq_documentation_block:
; Sequence 4, Application US/09103429A
; Patent No. 6187558
; GENERAL INFORMATION:
; APPLICANT: Granados, Robert R
; APPLICANT: Wang, Ping
; TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
; TITLE OF INVENTION: cDNA and Related Products and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
; STREET: 118 No. 6187558th f10ga
; CITY: Ithaca
; STATE: NY
; COUNTRY: USA
; ZIP: 14850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,429A
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39

; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Michaels, Christopher A
; REGISTRATION NUMBER: 34,390
; REFERENCE/DOCKET NUMBER: BTI-39
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (607) 256-2000
; TELEFAX: (607) 256-3628
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 805 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Trichoplusia ni
; TISSUE TYPE: peritrophic membrane
; US-09-103-429A-4

alignment_scores:
    Quality: 59.00      Length: 28
    Ratio: 3.933       Gaps: 1
    Percent Similarity: 53.571      Percent Identity: 46.429

alignment_block:
US-09-049-696-15 x US-09-103-429A-4
..
Align seg 1/1 to: US-09-103-429A-4 from: 1 to: 805

67 GCACGAGTATCTTTGTTATTCCTCCACAGACTCCGCCAGACAGA..... 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
507 AlaAlaProSerThrValValProAlaThrProAlaThrAlaAl 523

112 .CCTAGTCTGTGAACGTCGCTCCTGTGCT 144
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
523 aProValProThrThrAlaIleProThrPro 534

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-252-966B-12

seq_documentation_block:
; Sequence 12, Application US/08252966B
; Patent No. 5624818
; GENERAL INFORMATION:
; APPLICANT: Eisenman, Robert N.
; APPLICANT: Hurlin, Peter J.
; APPLICANT: Aver, Donald E.
; TITLE OF INVENTION: Regulatory proteins that dimerize with
; TITLE OF INVENTION: Mad or Max
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson, and KindnessPLLC
; STREET: 1420 Fifth Ave., Suite 2800
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/252,966B
; FILING DATE: 01-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997.
; REFERENCE/DOCKET NUMBER: FHCR17694
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-8100
; TELEFAX: (206) 224-0779
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:

```



```

; LENGTH: 1253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: translation of msina cDNA; see Figure 23
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
;
US-08-252-966B-12
..
alignment_scores:
    Quality: 58.50      Length: 47
    Ratio: 2.017        Gaps: 2
Percent Similarity: 61.702 Percent Identity: 34.043
alignment_block:
US-09-049-696-15 x US-08-252-966B-12 ..
Align seg 1/1 to: US-08-252-966B-12 from: 1 to: 1253
13 ATTGCTATTTCAGGCTGTTGATAGTCGATCGAATCA.....GAAAT 56
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
210 lleGluValGlnThrAsnAspMetValAsnValThrProGlyGlnVa 226
57 ATCCAAACATTGCACGAGTATCTTTGTTTATTCCTCCACAGACTCCGCCAG 106
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
226 lHisGlnIleProThrHisGlyIleGlnProGlnProGlnProProG 243
107 AGACACCTAGT...CCTGATGAACGTCGTCTGCTGTGCT 144
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
243 InHisProSerGlnProSerSerGlnSerAlaProThrPro 256
seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-252-966B-18

seq_documentation_block:
; Sequence 18, Application US/08252966B
; Patent No. 5624818
; GENERAL INFORMATION:
; APPLICANT: Eisenman, Robert N.
; APPLICANT: Hurlin, Peter J.
; APPLICANT: Ayer, Donald E.
; TITLE OF INVENTION: Regulatory Proteins that Dimerize with
; TITLE OF INVENTION: Mad or Max
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson, and KindnessPLLC
; STREET: 1420 Fifth Ave., Suite 2800
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/252,966B
; FILING DATE: 01-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997.
; REFERENCE/DOCKET NUMBER: FHCRI7694
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-8100
; TELEFAX: (206) 224-0779
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1261 amino acids
; TYPE: amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: translation of msina9 cDNA; see Figure 29A, B, C, D
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Mus musculus
;
US-08-252-966B-18
..
alignment_scores:
    Quality: 58.50      Length: 47
    Ratio: 2.017        Gaps: 2
Percent Similarity: 61.702 Percent Identity: 34.043
alignment_block:
US-09-049-696-15 x US-08-252-966B-18 ..
Align seg 1/1 to: US-08-252-966B-18 from: 1 to: 1261
13 ATTGCTATTTCAGGCTGTTGATAGTCGATCGAATCA.....GAAAT 56
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
210 lleGluValGlnThrAsnAspMetValAsnValThrProGlyGlnVa 226
57 ATCCAAACATTGCACGAGTATCTTTGTTTATTCCTCCACAGACTCCGCCAG 106
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
226 lHisGlnIleProThrHisGlyIleGlnProGlnProGlnProProG 243
107 AGACACCTAGT...CCTGATGAACGTCGTCTGCTGTGCT 144
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
243 InHisProSerGlnProSerSerGlnSerAlaProThrPro 256
seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-922-635-22

seq_documentation_block:
; Sequence 22, Application US/08922635A
; Patent No. 6033871
; GENERAL INFORMATION:
; APPLICANT: PILETZ, John E.
; APPLICANT: IVANOV, Tina R.
; TITLE OF INVENTION: DNA MOLECULES ENCODING IMIDALINE RECEPTIVE POLYPEPTIDES
; TITLE OF INVENTION: AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: Corrected Sequence Listing
; Patent No. 6033871
; CURRENT APPLICATION NUMBER: US/08/922,635A
; CURRENT FILING DATE: 1997-09-03
; EARLIER APPLICATION NUMBER: 08/650,766
; EARLIER FILING DATE: 1996-05-20
; EARLIER APPLICATION NUMBER: 60/012,600
; EARLIER FILING DATE: 1996-03-01
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1070
; TYPE: PRT
; ORGANISM: Homo sapiens
;
US-08-922-635-22
..
alignment_scores:
    Quality: 58.00      Length: 31
    Ratio: 2.636        Gaps: 1
Percent Similarity: 70.968 Percent Identity: 38.710
alignment_block:
US-09-049-696-15 x US-08-922-635-22 ..
Align seg 1/1 to: US-08-922-635-22 from: 1 to: 1070
67 GCACGAGTATCTTTGTTTATTCCTCCACAGACTCCGCCAGACACCTAG 116
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64g AlaSerThrSerAlaLeuValProGluGluThrProValGluAlaProAl 662

```


Align seg 1/1 to: AAG73854 from: 1 to: 552

1 ACAGATCTTTTCATGCTATTTCAGGCTGTGATAAGTTCGATCTGAAATC 50
|||||
476 ThrAspLeuPheileAlaalleGlnAlaValAspLysValAspLeuLys 492
51 AGAAATATCCAAATGCGACGAGTATCTTTGTTTATCTCTCCACAGATC 100
|||||
492 rGluLeuSerAsnIleAlaArgValSerLeuPheileProProGlnThrP 509
101 CGCCAGAGACACCTAGTCTGTGATGAACGCTGCTCTCTGCTTAATATT 150
|||||
509 roProGluThrProSerProAspGluThrSerAlaProCysProAsnIle 525

151 CAT 153
|||
526 His 526

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAG75614

seq_documentation_block:

ID AAG75614 standard; Protein; 869 AA.

XX AAG75614;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:6378.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma.

XX Homo sapiens.

XX WO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

XX 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI; 2001-235357/24.

XX N-PSDB; AAH35019.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -

XX Claim 11; Page 7851-7854; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAG77789 represent sequences used in the exemplification of the
XX present invention.
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were

CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 869 AA;

alignment_scores:
Quality: 265.00 Length: 51
Ratio: 5.196 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-15 x AAG75614

Align seg 1/1 to: AAG75614 from: 1 to: 869

1 ACAGATCTTTTCATGCTATTTCAGGCTGTGATAAGTTCGATCTGAAATC 50
|||||
793 ThrAspLeuPheileAlaalleGlnAlaValAspLysValAspLeuLys 809
51 AGAAATATCCAAATGCGACGAGTATCTTTGTTTATCTCTCCACAGATC 100
|||||
809 rGluLeuSerAsnIleAlaArgValSerLeuPheileProProGlnThrP 826
101 CGCCAGAGACACCTAGTCTGTGATGAACGCTGCTCTCTGCTTAATATT 150
|||||
826 roProGluThrProSerProAspGluThrSerAlaProCysProAsnIle 842

151 CAT 153
|||
843 His 843

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT: AAB74824

seq_documentation_block:

ID AAB74824 standard; Protein; 914 AA.

XX AAB74824;

XX 13-JUN-2001 (first entry)

XX Human ICACC-1 protein sequence.

XX ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
XX interleukin 9 induced calcium activated chloride channel; IL-9;
XX calcium activated chloride channel; anti-allergic; anti-asthmatic;
XX anti-inflammatory; immunomodulatory; cystic fibrosis;
XX inflammatory bowel disease; autoimmune disease.

XX Homo sapiens.

XX WC9944620-A1.

XX 10-SEP-1999.

XX 03-MAR-1999; 99WO-US04703.

XX 03-MAR-1998; 98US-0076815.

XX (MAGA-) MAGATININ PHARM INC.

XX Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;
XX Nicolaides NC, Zhou Y, Dong Q;

XX WPI; 1999-550979/46.

XX N-PSDB; AAF81927.

XX New nucleic acid encoding calcium activated chloride channel, used to
XX identify, e.g. specific modulators for treating atopic allergy -

XX Claim 11; Fig 4B; 75pp; English.

XX The present sequence represents the human interleukin 9 (IL-9) induced

CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
 CC have anti-allergic, anti-asthmatic, anti-inflammatory and
 CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
 CC used to alleviate asthma (or more generally atopic allergy), while those
 CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
 CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
 CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
 CC (increased levels) or IBD (reduced levels), also for monitoring
 CC treatment of these conditions. The ICACC proteins can be used:
 CC (i) to raise specific antibodies (Ab), useful: (a) as immunoassay
 CC reagents, and (b) as therapeutic (A); (ii) as ligands; and (iii) to
 CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
 CC identify modulators and binding partners. ICACC polynucleotides can be
 CC used to generate transgenic animals or recombinant cells, used to screen
 CC for antagonists, also as a source of therapeutic antisense agents or
 CC diagnostic probes (for quantifying mRNA expression, e.g. for
 CC identification of modulators).

XX Sequence 914 AA;

alignment_scores:
 Quality: 265.00 Length: 51
 Ratio: 5.196 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-15 x AAB74824 ..

Align seg 1/1 to: AAB74824 from: 1 to: 914

1 ACAGATCTTTTCATTGCTATTTCAGGCTGTTGATAGGTCGATCTCAATC 50
 |||||
 838 ThrAspLeuPheilleAlaileGlnAlaValAspLysValAspLeuLysSe 854
 51 AGAATATCCAAACATTGCACGAGTATCTTTGTTATTCTCCACAGACTC 100
 |||||
 854 rGluleSerAsnilleAlaArgValSerLeuPheilleProGlnThrp 871
 101 GCCCAGACAGACCTAGTCTCTGATGAACGCTCTGCTCTTGCTTAATTT 150
 |||||
 871 roProGluThrProSerProAspGluThrSerAlaProCysProAsnille 887

151 CAT 153

|||

888 His 888

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA24514

seq_documentation_block:

ID: AA24514 standard; Protein: 914 AA.

XX AA24514;

XX 12-OCT-2001 (first entry)

DE C902P predicted amino acid sequence.

XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;
 KW immunogenic; gene therapy; vaccine; colonic cancer.

XX Homo sapiens.

XX WO200149716-A2.

XX 12-JUL-2001.

XX 29-DEC-2000; 2000WO-US35596.

XX 30-DEC-1999; 99US-0476296.

PR 10-JAN-2000; 2000US-0480321.

PR 15-FEB-2000; 2000US-0504629.

PR 06-MAR-2000; 2000US-0519444.

PR 19-MAY-2000; 2000US-0575251.

PR 29-JUN-2000; 2000US-0609448.

XX 28-AUG-2000; 2000US-0649811.

PA (CORI-) CORIXA CORP.

XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;

PI King GE, Wang T, Jiang Y;

XX WPI; 2001-441847/47.

XX Colon tumor associated proteins and nucleic acids useful for the
 PT prevention, diagnosis and treatment of colonic cancer -

PS Claim 2; Page 440-443; 472pp; English.

XX The present invention describes colon tumour associated proteins (I) and
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumour associated protein (TCAP)
 CC expression, such as colonic cancer. For example, (I) and (II) may be
 CC used to treat disorders associated with decreased expression by
 CC rectifying mutations or deletions in a patient's genome that affect the
 CC activity of TCAPs by expressing inactive proteins or to supplement the
 CC patients own production of them. Additionally, (II) may be used to
 CC produce the TCAP proteins, by inserting the nucleic acids into a host
 CC cell culturing the cell to express the protein. (II) and its
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and
 CC quantitate the presence of similar nucleic acids in samples, and
 CC therefore which patients may be in need of restorative therapy. (I) may
 CC also be used as antigens in the production of antibodies against TCAPs
 CC and in assays to identify modulators of TCAP expression and activity.
 CC Anti-(I) antibodies and antagonists may also be used to down regulate
 CC TCAP expression and activity. The anti-(I) antibodies may also be used
 CC as diagnostic agents for detecting the presence of TCAPs in samples
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)). AA128460 to AA129512
 CC and AA24494 to AA24523 represent nucleotide and amino acid sequences
 CC given in the exemplification of the present invention.

XX Sequence 914 AA;

alignment_scores:

Quality: 265.00 Length: 51
 Ratio: 5.196 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-15 x AA24514 ..

Align seg 1/1 to: AA24514 from: 1 to: 914

1 ACAGATCTTTTCATTGCTATTTCAGGCTGTTGATAGGTCGATCTCAATC 50
 |||||
 838 ThrAspLeuPheilleAlaileGlnAlaValAspLysValAspLeuLysSe 854

51 AGAATATCCAAACATTGCACGAGTATCTTTGTTATTCTCCACAGACTC 100

|||||

854 rGluleSerAsnilleAlaArgValSerLeuPheilleProGlnThrp 871

101 GCCCAGACAGACCTAGTCTCTGATGAACGCTCTGCTCTTGCTTAATTT 150

|||||

871 roProGluThrProSerProAspGluThrSerAlaProCysProAsnille 887

151 CAT 153

|||

888 His 888

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AA24514

seq_documentation_block:

AD AAB73716 standard; Protein; 914 AA.
AC AAB73716;
XX
XX
XX 11-SEP-2001 (first entry)
XX
XX Human CLCA1 protein, SEQ ID NO:2.
DE
DE
XX Human CLCA1; goblet cell; mouse Gob-5 orthologue; drug screening;
KW expression inhibition; antisense therapy; gene therapy;
KW chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 152 /note= "Encoded by AGG in AAH46124"
FT
XX
XX WO200138530-A1.
XX
PD 31-MAY-2001.
XX
XX 22-NOV-2000; 2000WO-JP08232.
XX
XX 24-NOV-1999; 99JP-0333479.
PR 27-APR-2000; 2000JP-0127589.
XX
XX (TAKE) TAKEDA CHEM IND LTD.
XX
XX
XX Nakanishi A, Morita S;
PI
XX
XX WPI; 2001-355935/37.
DR N-PSDB; AAH46102, AAH46124.
XX
XX New antisense nucleotide, useful for treatment and prevention of
PT bronchial asthma and chronic obstructive pulmonary disease -
XX
XX
PS Claim 2: Page 76-80; 104pp; Japanese.
XX
XX The invention relates to an antisense nucleotide targetted to the mouse
CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
CC the CLCA1 gene (coding sequence shown in AAH46102). The invention also
CC relates to an antibody specific for the Gob-5 protein, medical and
CC diagnostic compositions containing the antisense nucleotide or the
CC antibody, and methods and kits for screening for compounds which inhibit
CC the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.
CC The antisense oligonucleotides and antibody are therefore useful for the
CC treatment and prevention of bronchial asthma and chronic obstructive
CC pulmonary disease. The present sequence represents human CLCA1 protein.
XX
XX
SQ Sequence 914 AA;

alignment_scores:
Quality: 265.00 Length: 51
Ratio: 5.196 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-049-696-15 x AAB73716 ..
Align seg 1/1 to: AAB73716 from: 1 to: 914
1 ACAGATCTTTCATTGCTATTCAGGCTGTTGATAGGTCGATCTGAATC 50
838 ThrAspLeuPheIleAlaIleAlaIleAlaValAspLeuLysSe 854
51 AGAATATCCACATTCGACGAGTATCTTTGTTTATTCCTCCACAGACT 100
854 rGluLeuSerAsnIleAlaArgValSerLeuPheIleProGlnThrP 871
101 GCCAGACAGACCTAGTCTCTGATGAACGCTGCTCTCTGCTTAATATT 150
|||||

871 roProGluThrProSerProAspGluThrSerAlaProCysProAsnIle 887
151 CAT 153
888 Hls 888
seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAB74733
seq_documentation_block:
ID: AAB74733 standard; Protein; 914 AA.
XX
AC AAB74733;
XX
XX 12-JUN-2001 (first entry)
XX
XX Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.
XX
KW Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW immunostimulant; cytostatic; cardiac; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; neurotropic; anticonvulsant; vaccine;
KW antialzheimer; antiparkinsonian; antimicrobial; vulnery; gene therapy;
KW immune disorder; hyperproliferative disorder; cardiovascular disease;
KW cancer; angiogenic disorder; neurological disorder; infectious disease;
KW wound healing; regeneration; chemotaxis; chromosome 1.
XX
XX Homo sapiens.
XX
XX WO200112775-A2.
XX
XX 22-FEB-2001.
XX
XX 16-AUG-2000; 2000WO-US22325.
XX
XX 17-AUG-1999; 99US-0149182.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX
XX Rosen CA, Ni J, Florence KA, Fiscella M, Wei P, Baker KP;
XX Birse CE, Young PE, Komatsuoulls GA, Moore PA, Soppet DR;
XX
XX WPI; 2001-147550/15.
XX
XX N-PSDB; AAF81787.
XX
XX Nucleic acids encoding 25 human secreted polypeptides, useful for
XX preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX disease and diabetic retinopathy -
XX
XX
XX Claim 11; Page 459-460; 485pp; English.
XX
XX AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733
XX to AAB74772. Human secreted proteins can have activities based on the
XX tissues and cells they are expressed in. Example of activities include:
XX immunomodulatory; antisclerotic; dermatological; immunosuppressive;
XX antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiac;
XX vascular; anti-angiogenic; ophthalmological; neuroprotectant; neurotropic;
XX anticonvulsant; antialzheimer; antiparkinsonian; antimicrobial; and
XX vulnery. Human secreted proteins can be used in gene therapy and
XX vaccine. Human secreted protein nucleotide sequences (NAMI) and proteins
XX (PEPI) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate polypeptide expression. For example, NAMI
XX and PEPI may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patients genome
XX that affect the activity of proteins by expressing inactive proteins or
XX to supplement the patients own production of polypeptides. Disorders that
XX may be prevented, diagnosed and/or treated include immune disorders,
XX hyperproliferative disorders (e.g. cancers), cardiovascular diseases,
XX angiogenic disorders, neurological disorders, infectious diseases and/or
XX for promoting wound healing, regeneration and /or chemotaxis. AAF81778 to
XX AAF81786 and AAB74732 represent sequences used in the exemplification of
XX the present invention.
XX
XX
SQ Sequence 914 AA;

alignment_scores: Quality: 265.00 Length: 51
Ratio: 5.196 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-049-696-15 x AAB74733 ..
Align seg 1/1 to: AAB74733 from: 1 to: 914

1 ACAGATCTTTTCATGCTATTCAGGCTGTTGATAAGGTCGATCTGAAATC 50
838 ThrAspLeuPheIleAlaIleGlnAlaValAspLysValAspLeuLys 854
51 AGAATATCCAAACATTCACGAGTATCTTTGTTTATTCTCCACAGACTC 100
854 rGlulSerAsnIleAlaArgValSerLeuPheIleProGlnThrP 871
101 CGCCAGAGACACCTAGTCTGCTGATGAACGCTGCTCTGCTTGTCTTAAT 150
871 roProGluThrProSerProAspGluThrSerAlaProCysProAsnIle 887
151 CAT 153
888 His 888

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAAG75474

seq_documentation_block:
ID AAG75474 standard; Protein; 925 AA.
XX AC AAG75474;
XX AC (first entry)
XX 03-SEP-2001 (first entry)
XX Human colon cancer antigen protein SEQ ID NO:6238.
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 1.
XX Homo sapiens.
XX WO200122920-A2.
XX 05-APR-2001.
XX 28-SEP-2000; 2000WO-US26524.
XX 29-SEP-1999; 99US-0157137.
XX 03-NOV-1999; 99US-0163280.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
XX N-PSDB; AAH34879.

Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
useful for preventing, diagnosing and/or treating colorectal cancers -
Claim 11; Page 7686-7690; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions

in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally; N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAB7789 represent sequences used in the exemplification of the
XX present invention.
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.

Sequence 925 AA;
alignment_scores: Quality: 265.00 Length: 51
Ratio: 5.196 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-049-696-15 x AAG75474 ..
Align seg 1/1 to: AAG75474 from: 1 to: 925
1 ACAGATCTTTTCATGCTATTCAGGCTGTTGATAAGGTCGATCTGAAATC 50
849 ThrAspLeuPheIleAlaIleGlnAlaValAspLysValAspLeuLys 865
51 AGAATATCCAAACATTCACGAGTATCTTTGTTTATTCTCCACAGACTC 100
865 rGlulSerAsnIleAlaArgValSerLeuPheIleProGlnThrP 882
101 CGCCAGAGACACCTAGTCTGCTGATGAACGCTGCTCTGCTTGTCTTAAT 150
882 roProGluThrProSerProAspGluThrSerAlaProCysProAsnIle 898
151 CAT 153
899 His 899

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT:AAW06548
seq_documentation_block:
ID AAW06548 standard; Protein; 228 AA.
XX AC AAW06548;
XX 13-MAR-1997 (first entry)
XX Human colon specific gene CSG5 polypeptide fragment.
XX Colon specific gene; CSG5; colon cancer; metastasis; diagnosis;
XX therapy; antibody; vaccine.
XX Homo sapiens.
XX WO9639419-A1.
XX 12-DEC-1996.
XX 06-JUN-1995; 95WO-US07289.
XX 06-JUN-1995; 95WO-US07289.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Yu G;
XX WPI; 1997-043054/04.
XX N-PSDB; AAH45884.
XX human colon specific genes and their expression products - detection

PT of which, in non-colon tissue samples, can be used as indication of
XX colon cancer metastasis

PS Claim 8; Fig 5; 60pp; English.

XX Novel polypeptides (AAW06545-53) are encoded by cDNA clones (see also
CC AAW4580-92) corresponding to 13 human colon specific genes,
CC designated CSG1, CSG2, etc., that are primarily expressed in
CC tissues derived from the colon. Recombinant CSG polypeptides can
CC be produced in transformed host cells. They are useful diagnostic
CC markers for colon cancer and for colon cancer metastasis and can
CC also be used to screen for (ant)agonist cpds. of therapeutic or
CC diagnostic value. Antibodies raised against the colon-specific
CC polypeptides may be used to target colon cancer cells or as part
CC of a colon cancer vaccine.

XX Sequence 228 AA;

alignment_scores:
Quality: 240.00 Length: 47
Ratio: 5.106 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-15 x AAW06548 ..

Align seg 1/1 to: AAW06548 from: 1 to: 228

1 ACAGATCTTTTCATTGCTATTTCAGGCTGTGATAGCTCGATCTGAAATC 50
|||||
182 ThrAspLeuPheIleAlaIleGlnAlaValAspLysValAspLeuLysSe 198

51 AGAAATATCCAACTTCGACGAGTATCTTGTATTATTCCTCCACAGACTC 100
|||||
198 rGluIleSerAsnIleAlaArgValSerLeuPheIleProGlnThrp 215

101 CGCCAGACACCTAGTCTCTGATGAACGTCCTCTCTTGT 141
|||||

215 roProGluThrProSerProAspGluThrSerAlaProCys 228

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW46879

seq_documentation_block:

ID AAW46879 standard; Protein: 228 AA.

AC AAW46879;

DT 22-JUN-1998 (first entry)

DE Protein sequence encoded by a colon-specific gene.

XX Colon-specific gene; probe; detection; expression; human;
KW diagnostic assay; colon cancer; antibody; screening.

OS Homo sapiens.

PN US5733748-A.

PD 31-MAR-1998.

PF 06-JUN-1995; 95US-0469667.

PR 06-JUN-1995; 95US-0469667.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen C, Yu G;

XX WPI; 1998-229823/20.

XX N-PSDB; AAV16672.

XX Colon-specific nucleic acids - useful as probes for detecting colon

PT cancer micrometastases

XX Claim 1; Fig 5; 51pp; English.

XX AAW46876-80 and W4682-85 represent proteins encoded by colon-specific
CC genes. The polynucleotides encoding these proteins can be used
CC as probes to detect expression of the corresponding human genes,
CC e.g. in diagnostic assays for detecting micrometastases of colon cancer.
CC Recombinant cells containing the polynucleotides can be used to
CC produce the proteins, in order that antibodies can be raised and
CC used in further screening or diagnostics.

XX Sequence 228 AA;

alignment_scores:
Quality: 240.00 Length: 47
Ratio: 5.106 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-15 x AAW46879 ..

Align seg 1/1 to: AAW46879 from: 1 to: 228

1 ACAGATCTTTTCATTGCTATTTCAGGCTGTTCATAAGCTCGATCTGAAATC 50
|||||
182 ThrAspLeuPheIleAlaIleGlnAlaValAspLysValAspLeuLysSe 198

51 AGAAATATCCAACTTCGACGAGTATCTTGTATTATTCCTCCACAGACTC 100
|||||
198 rGluIleSerAsnIleAlaArgValSerLeuPheIleProGlnThrp 215

101 CGCCAGACACCTAGTCTCTGATGAACGTCCTCTCTTGT 141
|||||

215 roProGluThrProSerProAspGluThrSerAlaProCys 228

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW4822

seq_documentation_block:

ID AAW4822 standard; Protein: 913 AA.

AC AAW4822;

DT 13-JUN-2001 (first entry)

DE Define ICACC-1 protein sequence.

XX ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW inflammatory bowel disease; autoimmune disease.

XX Mus sp.

PN WC9944620-A1.

PD 10-SEP-1999.

PF 03-MAR-1999; 99WO-US04703.

PR 03-MAR-1998; 98US-0076815.

XX (MAGA-) MAGAININ PHARM INC.

XX Hoiroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;

XX Nicolaides NC, Zhou Y, Dong Q;

XX WPI; 1999-550979/46.

XX N-PSDB; AAF81925.

XX New nucleic acid encoding calcium activated chloride channel, used to

PR 02-JUN-1998; 98US-0087759.
PR 03-JUN-1998; 98US-0087827.
PR 04-JUN-1998; 98US-0088021.
PR 05-JUN-1998; 98US-0088025.
PR 06-JUN-1998; 98US-0088028.
PR 07-JUN-1998; 98US-0088029.
PR 08-JUN-1998; 98US-0088030.
PR 09-JUN-1998; 98US-0088033.
PR 10-JUN-1998; 98US-0088036.
PR 11-JUN-1998; 98US-0088167.
PR 12-JUN-1998; 98US-0088202.
PR 13-JUN-1998; 98US-0088212.
PR 14-JUN-1998; 98US-0088217.
PR 15-JUN-1998; 98US-0088655.
PR 16-JUN-1998; 98US-0088722.
PR 17-JUN-1998; 98US-0088730.
PR 18-JUN-1998; 98US-0088734.
PR 19-JUN-1998; 98US-0088738.
PR 20-JUN-1998; 98US-0088740.
PR 21-JUN-1998; 98US-0088741.
PR 22-JUN-1998; 98US-0088742.
PR 23-JUN-1998; 98US-0088810.
PR 24-JUN-1998; 98US-0088811.
PR 25-JUN-1998; 98US-0088824.
PR 26-JUN-1998; 98US-0088825.
PR 27-JUN-1998; 98US-0088826.
PR 28-JUN-1998; 98US-0088858.
PR 29-JUN-1998; 98US-0088861.
PR 30-JUN-1998; 98US-0088863.
PR 31-JUN-1998; 98US-0088876.
PR 01-JUL-1998; 98US-0089090.
PR 02-JUL-1998; 98US-0089105.
PR 03-JUL-1998; 98US-0089440.
PR 04-JUL-1998; 98US-0089512.
PR 05-JUL-1998; 98US-0089514.
PR 06-JUL-1998; 98US-0089532.
PR 07-JUL-1998; 98US-0089538.
PR 08-JUL-1998; 98US-0089598.
PR 09-JUL-1998; 98US-0089599.
PR 10-JUL-1998; 98US-0089600.
PR 11-JUL-1998; 98US-0089653.
PR 12-JUL-1998; 98US-0089801.
PR 13-JUL-1998; 98US-0089807.
PR 14-JUL-1998; 98US-0089808.
PR 15-JUL-1998; 98US-0089847.
PR 16-JUL-1998; 98US-0089948.
PR 17-JUL-1998; 98US-0089952.
PR 18-JUL-1998; 98US-0090246.
PR 19-JUL-1998; 98US-0090252.
PR 20-JUL-1998; 98US-0090254.
PR 21-JUL-1998; 98US-0090349.
PR 22-JUL-1998; 98US-0090355.
PR 23-JUL-1998; 98US-0090429.
PR 24-JUL-1998; 98US-0090431.
PR 25-JUL-1998; 98US-0090435.
PR 26-JUL-1998; 98US-0090444.
PR 27-JUL-1998; 98US-0090445.
PR 28-JUL-1998; 98US-0090461.
PR 29-JUL-1998; 98US-0090472.
PR 30-JUL-1998; 98US-0090535.
PR 31-JUL-1998; 98US-0090538.
PR 01-AUG-1998; 98US-0090540.
PR 02-AUG-1998; 98US-0090557.
PR 03-AUG-1998; 98US-0090676.
PR 04-AUG-1998; 98US-0090678.
PR 05-AUG-1998; 98US-0090688.
PR 06-AUG-1998; 98US-0090690.
PR 07-AUG-1998; 98US-0090691.
PR 08-AUG-1998; 98US-0090694.
PR 09-AUG-1998; 98US-0090695.
PR 10-AUG-1998; 98US-0090896.
PR 11-AUG-1998; 98US-0090862.
PR 12-AUG-1998; 98US-0090863.

PR 01-JUL-1998; 98US-0091358.
PR 02-JUL-1998; 98US-0091360.
PR 03-JUL-1998; 98US-0091544.
PR 04-JUL-1998; 98US-0091478.
PR 05-JUL-1998; 98US-0091486.
PR 06-JUL-1998; 98US-0091519.
PR 07-JUL-1998; 98US-0091626.
PR 08-JUL-1998; 98US-0091628.
PR 09-JUL-1998; 98US-0091633.
PR 10-JUL-1998; 98US-0091646.
PR 11-JUL-1998; 98US-0091673.
PR 12-JUL-1998; 98US-0091978.
PR 13-JUL-1998; 98US-0091982.
PR 14-JUL-1998; 98US-0092182.
PR 15-JUL-1998; 98US-0092472.
PR 16-JUL-1998; 98US-0093339.
PR 17-JUL-1998; 98US-0094651.
PR 18-JUL-1998; 98US-0095282.
PR 19-JUL-1998; 98US-0095285.
PR 20-JUL-1998; 98US-0095301.
PR 21-JUL-1998; 98US-0095302.
PR 22-JUL-1998; 98US-0095318.
PR 23-JUL-1998; 98US-0095321.
PR 24-JUL-1998; 98US-0095325.
PR 25-JUL-1998; 98US-0095916.
PR 26-JUL-1998; 98US-0095929.
PR 27-JUL-1998; 98US-0096012.
PR 28-JUL-1998; 98US-0096143.
PR 29-JUL-1998; 98US-0096146.
PR 30-JUL-1998; 98US-0096329.
PR 31-JUL-1998; 98US-0096757.
PR 01-AUG-1998; 98US-0096766.
PR 02-AUG-1998; 98US-0096768.
PR 03-AUG-1998; 98US-0096773.
PR 04-AUG-1998; 98US-0096791.
PR 05-AUG-1998; 98US-0096867.
PR 06-AUG-1998; 98US-0096891.
PR 07-AUG-1998; 98US-0096894.
PR 08-AUG-1998; 98US-0096895.
PR 09-AUG-1998; 98US-0096897.
PR 10-AUG-1998; 98US-0096949.
PR 11-AUG-1998; 98US-0096950.
PR 12-AUG-1998; 98US-0096959.
PR 13-AUG-1998; 98US-0096960.
PR 14-AUG-1998; 98US-0097022.
PR 15-AUG-1998; 98US-0097141.
PR 16-AUG-1998; 98US-0097218.
PR 17-AUG-1998; 98US-0097661.
PR 18-AUG-1998; 98US-0097951.
PR 19-AUG-1998; 98US-0097952.
PR 20-AUG-1998; 98US-0097954.
PR 21-AUG-1998; 98US-0097955.
PR 22-AUG-1998; 98US-0097971.
PR 23-AUG-1998; 98US-0097974.
PR 24-AUG-1998; 98US-0097978.
PR 25-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 27-AUG-1998; 98US-0098014.
PR 28-AUG-1998; 98US-0098525.
PR 29-AUG-1998; 98US-0100634.
PR 30-AUG-1998; 98US-0100634.
PR 31-AUG-1998; 98US-0100634.
PR 01-SEP-1998; 98US-0100634.
PR 02-SEP-1998; 98US-0100634.
PR 03-SEP-1998; 98US-0100634.
PR 04-SEP-1998; 98US-0100634.
PR 05-SEP-1998; 98US-0100634.
PR 06-SEP-1998; 98US-0100634.
PR 07-SEP-1998; 98US-0100634.
PR 08-SEP-1998; 98US-0100634.
PR 09-SEP-1998; 98US-0100634.
PR 10-SEP-1998; 98US-0100634.
PR 11-SEP-1998; 98US-0100634.
PR 12-SEP-1998; 98US-0100634.
PR 13-SEP-1998; 98US-0100634.
PR 14-SEP-1998; 98US-0100634.
PR 15-SEP-1998; 98US-0100634.
PR 16-SEP-1998; 98US-0100634.
PR 17-SEP-1998; 98US-0100634.
PR 18-SEP-1998; 98US-0100634.
PR 19-SEP-1998; 98US-0100634.
PR 20-SEP-1998; 98US-0100634.
PR 21-SEP-1998; 98US-0100634.
PR 22-SEP-1998; 98US-0100634.
PR 23-SEP-1998; 98US-0100634.
PR 24-SEP-1998; 98US-0100634.
PR 25-SEP-1998; 98US-0100634.
PR 26-SEP-1998; 98US-0100634.
PR 27-SEP-1998; 98US-0100634.
PR 28-SEP-1998; 98US-0100634.
PR 29-SEP-1998; 98US-0100634.
PR 30-SEP-1998; 98US-0100634.
PR 01-OCT-1998; 98US-0100634.
PR 02-OCT-1998; 98US-0100634.
PR 03-OCT-1998; 98US-0100634.
PR 04-OCT-1998; 98US-0100634.
PR 05-OCT-1998; 98US-0100634.
PR 06-OCT-1998; 98US-0100634.
PR 07-OCT-1998; 98US-0100634.
PR 08-OCT-1998; 98US-0100634.
PR 09-OCT-1998; 98US-0100634.
PR 10-OCT-1998; 98US-0100634.
PR 11-OCT-1998; 98US-0100634.
PR 12-OCT-1998; 98US-0100634.
PR 13-OCT-1998; 98US-0100634.
PR 14-OCT-1998; 98US-0100634.
PR 15-OCT-1998; 98US-0100634.
PR 16-OCT-1998; 98US-0100634.
PR 17-OCT-1998; 98US-0100634.
PR 18-OCT-1998; 98US-0100634.
PR 19-OCT-1998; 98US-0100634.
PR 20-OCT-1998; 98US-0100634.
PR 21-OCT-1998; 98US-0100634.
PR 22-OCT-1998; 98US-0100634.
PR 23-OCT-1998; 98US-0100634.
PR 24-OCT-1998; 98US-0100634.
PR 25-OCT-1998; 98US-0100634.
PR 26-OCT-1998; 98US-0100634.
PR 27-OCT-1998; 98US-0100634.
PR 28-OCT-1998; 98US-0100634.
PR 29-OCT-1998; 98US-0100634.
PR 30-OCT-1998; 98US-0100634.
PR 31-OCT-1998; 98US-0100634.
PR 01-NOV-1998; 98US-0100634.
PR 02-NOV-1998; 98US-0100634.
PR 03-NOV-1998; 98US-0100634.
PR 04-NOV-1998; 98US-0100634.
PR 05-NOV-1998; 98US-0100634.
PR 06-NOV-1998; 98US-0100634.
PR 07-NOV-1998; 98US-0100634.
PR 08-NOV-1998; 98US-0100634.
PR 09-NOV-1998; 98US-0100634.
PR 10-NOV-1998; 98US-0100634.
PR 11-NOV-1998; 98US-0100634.
PR 12-NOV-1998; 98US-0100634.
PR 13-NOV-1998; 98US-0100634.
PR 14-NOV-1998; 98US-0100634.
PR 15-NOV-1998; 98US-0100634.
PR 16-NOV-1998; 98US-0100634.
PR 17-NOV-1998; 98US-0100634.
PR 18-NOV-1998; 98US-0100634.
PR 19-NOV-1998; 98US-0100634.
PR 20-NOV-1998; 98US-0100634.
PR 21-NOV-1998; 98US-0100634.
PR 22-NOV-1998; 98US-0100634.
PR 23-NOV-1998; 98US-0100634.
PR 24-NOV-1998; 98US-0100634.
PR 25-NOV-1998; 98US-0100634.
PR 26-NOV-1998; 98US-0100634.
PR 27-NOV-1998; 98US-0100634.
PR 28-NOV-1998; 98US-0100634.
PR 29-NOV-1998; 98US-0100634.
PR 30-NOV-1998; 98US-0100634.
PR 01-DEC-1998; 98US-0100634.
PR 02-DEC-1998; 98US-0100634.
PR 03-DEC-1998; 98US-0100634.
PR 04-DEC-1998; 98US-0100634.
PR 05-DEC-1998; 98US-0100634.
PR 06-DEC-1998; 98US-0100634.
PR 07-DEC-1998; 98US-0100634.
PR 08-DEC-1998; 98US-0100634.
PR 09-DEC-1998; 98US-0100634.
PR 10-DEC-1998; 98US-0100634.
PR 11-DEC-1998; 98US-0100634.
PR 12-DEC-1998; 98US-0100634.
PR 13-DEC-1998; 98US-0100634.
PR 14-DEC-1998; 98US-0100634.
PR 15-DEC-1998; 98US-0100634.
PR 16-DEC-1998; 98US-0100634.
PR 17-DEC-1998; 98US-0100634.
PR 18-DEC-1998; 98US-0100634.
PR 19-DEC-1998; 98US-0100634.
PR 20-DEC-1998; 98US-0100634.
PR 21-DEC-1998; 98US-0100634.
PR 22-DEC-1998; 98US-0100634.
PR 23-DEC-1998; 98US-0100634.
PR 24-DEC-1998; 98US-0100634.
PR 25-DEC-1998; 98US-0100634.
PR 26-DEC-1998; 98US-0100634.
PR 27-DEC-1998; 98US-0100634.
PR 28-DEC-1998; 98US-0100634.
PR 29-DEC-1998; 98US-0100634.
PR 30-DEC-1998; 98US-0100634.
PR 31-DEC-1998; 98US-0100634.

(GETH) GENENTECH INC.
Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
Wang WL, Yuan J;
WPI; 2000-072883/06.
NPSDB; AAZ65095.
Membrane-bound proteins and related nucleotide sequences
claim 12; Fig 274; 822pp; English.

This Page Blank (uspto)

OM of: US-09-049-696-14 to: SPTREMBL_17.* out_format : pfs

Date: Mar 30, 2002 2:47 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+gfp.model -DEV=xlp
-Q/cgn2.1/uspro.spool/US09049696/runat_28032002_145238_2053/app_query.fasta_1.12579
-DB=SPTREMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-CAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcpt
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1_705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-14

Query length: 248

Database: SPTREMBL_17.*

Database sequences: 473505

Database length: 14627329

Search time (sec): 805.760000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
sp_human:Q95151	+	427.00	903.05	5.5e-43	914	! Q95151 homo sapiens (human). ca
sp_human:Q9UNF6	+	427.00	903.05	5.5e-43	914	! Q9unf6 homo sapiens (human). ca
sp_human:Q9UPC6	+	422.00	892.20	2.2e-42	914	! Q9upc6 homo sapiens (human). ca
sp_mammal:Q9TRB5	+	348.00	731.62	1.9e-33	917	! Q9trb5 sus scrofa (pig). epithe
sp_rodent:Q88826	+	304.00	636.20	4.0e-28	913	! Q88826 mus musculus (mouse). gc
sp_rodent:Q9D726	+	300.00	627.52	1.2e-27	913	! Q9d726 mus musculus (mouse). ch
sp_human:Q9NXP1	+	259.50	545.24	9.1e-23	469	! Q9nxp1 homo sapiens (human). co
sp_human:Q9UNF7	+	259.50	539.62	9.6e-23	917	! Q9unf7 homo sapiens (human). ca
sp_rodent:Q88860	+	210.50	433.46	8.0e-17	902	! Q88860 mus musculus (mouse). ch
sp_rodent:Q9OX15	+	210.50	433.46	8.0e-17	902	! Q9ox15 mus musculus (mouse). ch
sp_rodent:Q9E0R4	+	201.50	413.92	9.8e-16	902	! Q9e0r4 mus musculus (mouse). en
sp_rodent:Q9R070	+	198.50	407.41	2.3e-15	902	! Q9r070 mus musculus (mouse). ca
sp_mammal:O18741	+	159.00	399.79	6.0e-15	905	! O18741 bos taurus (bovine). lu
sp_human:Q9UQC9	+	159.00	321.34	1.3e-10	943	! Q9uqc9 homo sapiens (human). ch
sp_human:Q9Y6N2	+	159.00	321.34	1.3e-10	943	! Q9y6n2 homo sapiens (human). ca
sp_vertebrate:Q9W6V5	+	95.00	179.14	0.0075	1406	! Q9w6v5 gallus gallus (chicken)
sp_vertebrate:Q9R1L6	+	94.00	207.70	0.0075	36	! Q9r1l6 rattus norvegicus (rat).
sp_plant:Q9W7R5	-	77.50	148.11	0.9162	615	! Q9w7r5 odontella sinensis. tric
sp_plant:Q9LW22	+	73.00	139.48	3.17	537	! Q9ln22 arabidopsis thaliana (md
sp_bacteria:Q9RYC5	+	72.50	140.89	3.56	399	! Q9ryc5 deinococcus radiodurans
sp_organella:O47500	+	70.50	131.14	6.53	760	! O47500 venturia inaequalis. rt-
sp_virus:Q993M3	-	69.00	131.42	9.60	499	! Q993m3 autonomous rat parvoviru
sp_virus:P90329	-	69.00	130.13	9.71	582	! P90329 kilham rat virus. non-ca
sp_virus:Q993M4	-	69.00	129.34	9.78	639	! Q993m4 autonomous rat parvoviru
sp_virus:P88901	-	69.00	128.32	9.87	722	! P88901 kilham rat virus. non-ca
sp_invertebrate:Q9TW28	-	69.00	120.96	10.55	1737	! Q9tw28 dictyostelium discoid
sp_invertebrate:Q9NAF2	-	68.50	127.82	11.29	673	! Q9naf2 caenorhabditis elegans
sp_invertebrate:Q9VF16	+	68.00	139.22	11.59	152	! Q9vf16 drosophila melanogaste
sp_bacteria:Q24997	-	68.00	125.36	13.13	793	! Q24997 helicobacter pylori (cam
sp_virus:Q88496	-	68.00	124.12	13.28	920	! Q88496 theiler's encephalomyeli
sp_virus:Q88497	-	68.00	124.12	13.28	920	! Q88497 theiler's encephalomyeli
sp_vertebrate:P79757	-	68.00	123.24	13.39	1021	! P79757 gallus gallus (chicken)
sp_invertebrate:O17149	+	67.50	131.86	14.10	321	! O17149 caenorhabditis elegans
sp_invertebrate:Q9VBZ9	+	67.50	131.84	14.10	322	! Q9vzb9 drosophila melanogaste
sp_archaea:Q58743	-	67.50	131.48	14.15	336	! Q58743 methanococcus jannaschii
sp_plant:Q9SW36	+	67.50	129.55	14.39	423	! Q9sw36 arabidopsis thaliana (md
sp_invertebrate:O15724	+	66.50	129.67	18.62	322	! O15724 dictyostelium discoid
sp_fungi:Q94173	-	66.50	119.80	20.35	1044	! Q94173 pneumocystis carinii. v
sp_invertebrate:O00886	+	66.50	117.74	20.73	1335	! O00886 dictyostelium discoid
sp_archaea:O58693	+	66.00	129.20	21.28	299	! O58693 pyrococcus horikoshii. h

sp_virus:Q9W111 - 66.00 128.43 21.43 328 ! Q9w111 bovine herpesvirus 4.
sp_invertebrate:Q21554 - 66.00 118.03 23.54 1134 ! Q21554 caenorhabditis ele
sp_virus:Q9Q8Y6 - 65.50 122.25 25.79 602 ! Q9q8y6 human immunodeficienc
sp_invertebrate:Q9UIM9 - 65.50 118.69 26.63 920 ! Q9uim9 dictyostelium disco
sp_invertebrate:Q20036 + 65.00 130.41 27.27 200 ! Q20036 caenorhabditis eleg
seq_name: sp_human:Q95151

seq_documentation_block:

ID Q95151 PRELIMINARY; PRT; 914 AA.
AC Q95151;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC 1 TISSUE=SMALL INTESTINE;
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Eble R.C., Ji H.L., Schreuer K.D., Fuller C.M.,
Pauli B.O.;
RT 'Genomic cloning, molecular characterization, and functional analysis
of human HCLCAL, the first human member of the family of Ca2+-activated
Cl-channel proteins.';
EL Genomics 54:200-214(1998).
DR EMBL: AF039400; AAC95428.1; -
DR InterPro: IPR000131; ATPase_gamma.
DR InterPro: IPR020335; VWFA.
DR PROSITE: P500153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE: P50234; VWFA; 1.
DR SMART: SM00327; VWA; 1.
SQ SEQUENCE 914 AA; 100174 MW; 7066440C46557FA3 CRC64;

alignment_scores:

Quality: 427.00 Length: 82
Ratio: 5.207 Gaps: 0
Percent Similarity: 100.00 Percent Identity: 100.00

alignment_block:

US-09-049-696-14 x Q95151 ..
Align seg 1/1 to: Q95151 from: 1 to: 914

3 CTGAAGCGGAAATTCACGGGGGCGAGTCCTCATTAATCTGACGTGGACAGC 52
|||||
759 LeuLysAlaGluLeuHisGlyGlySerLeuLeuAsnLeuThrTrpThrAl 775
53 TCCCTGGGATCATTTATGACCATGGAACAGCTCACAGTATATCATTCGAA 102
|||||
775 aProGlyAspAspPyrAspHisGlyThrAlaHisLysThrIleLeuArgI 792
|||||
103 TAAGTACAAGTATTTCTTGATCTCAGACACAAGTTCATCAATGAACTCTCTCAA 152
|||||
792 LeSerThrSerIleLeuAspLeuArgAspLysPheAsnGluSerLeuGln 808
153 GTGAATACTACTGCTCTCATCCCCAAGAACCCAACTCTGAGGAAGTCTT 202
|||||
809 ValAsnThrThrAlaLeuLeuProLysGluAlaAsnSerGluGluValPh 825
203 TTTCTTTAAACCAACAACATTACTTTTGAATATGGCACACATCTT 248
|||||
825 eLeuPheLysProGluAsnIleThrPheGluAsnGlyThrAspLeu 840
seq_name: sp_human:Q9UNF6

seq_documentation_block:

ID Q9UNF6 PRELIMINARY; PRT; 914 AA.
AC Q9UNF6;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC 1 TISSUE=SMALL INTESTINE;
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Eble R.C., Ji H.L., Schreuer K.D., Fuller C.M.,
Pauli B.O.;
RT 'Genomic cloning, molecular characterization, and functional analysis
of human HCLCAL, the first human member of the family of Ca2+-activated
Cl-channel proteins.';
EL Genomics 54:200-214(1998).
DR EMBL: AF039400; AAC95428.1; -
DR InterPro: IPR000131; ATPase_gamma.
DR InterPro: IPR020335; VWFA.
DR PROSITE: P500153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE: P50234; VWFA; 1.
DR SMART: SM00327; VWA; 1.
SQ SEQUENCE 914 AA; 100174 MW; 7066440C46557FA3 CRC64;

AC Q9UNF6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 1.
GN CACCL
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE, AND COLON;
RX MEDLINE=99364503; PubMed=10437792;
RA Aguel M., Vermat T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
RT chloride channel (caccl) family predominantly expressed in the
RT digestive tract and trachea."
RL FEBS Lett. 455:295-301(1999).
DR EMBL; AF127036; AAD25487.1; .
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
SQ SEQUENCE 914 AA; 100202 MW; B58B451D97887D6B CRC64;

alignment_scores:
Quality: 427.00 Length: 82
Ratio: 5.207 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-14 x Q9UNF6 ..
Align seg 1/1 to: Q9UNF6 from: 1 to: 914

3 CTGAAGCGGAAATTCACGGGGCAGTCTCAATTAATCTGACTTGGACAGC 52
|||||
759 LeuLysAlaGluIleHisGlySerLeuIleAsnLeuThrTrpThrAl 775

53 TCCTGGGGATGATTATCACCAGTCAAGCTCACAAGTATATCATTCGAA 102
|||||
775 aProGlyAspPtyrAspHisGlyThrAlaHisLysTyrIleIleArgI 792

103 TAAGTACAAGTATTCTTGATCTCAGACAGCAAGTTCATCAATGAATCTCTCAA 152
|||||
792 leSerThrSerIleLeuAspLeuArgAspLysPheAsnGluSerLeuGln 808

153 GTGAATACTGCTGCTCATCCCAAGGAGCAACTCTGAGGAAGTCTT 202
|||||
809 ValAsnThrThrAlaLeuIleProLysGluAlaAsnSerGluGluValPh 825

203 TTTCTTTTAAACCAACAATTTACTTTTGAATAATGGCACAGATCTT 248
|||||
825 eLeuPheLysProGluAsnIleThrPheGluAsnGlyThrAspLeu 840

seq_name: sp_human:Q9UNF6

seq_documentation_block:
ID Q9UNF6 PRELIMINARY; PRT; 914 AA.
AC Q9UNF6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Elble R.C., Ji H.L., Schreur K.D., Fuller C.M.,
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
RT of human CLCAL, the first human member of the family of Ca2+-activated
RT Cl-channel proteins."
RL Genomics 54:200-214(1998).
DR EMBL; AF039401; AAC95429.1; .
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 914 AA; 100160 MW; 7761362B66752AA3 CRC64;

alignment_scores:
Quality: 422.00 Length: 82
Ratio: 5.146 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.780

alignment_block:
US-09-049-696-14 x Q9UPC6 ..
Align seg 1/1 to: Q9UPC6 from: 1 to: 914

3 CTGAAGCGGAAATTCACGGGGCAGTCTCAATTAATCTGACTTGGACAGC 52
|||||
759 LeuAsnAlaGluIleHisGlySerLeuIleAsnLeuThrTrpThrAl 775

53 TCCTGGGGATGATTATCACCAGTCAAGCTCACAAGTATATCATTCGAA 102
|||||
775 aProGlyAspPtyrAspHisGlyThrAlaHisLysTyrIleIleArgI 792

103 TAAGTACAAGTATTCTTGATCTCAGACAGCAAGTTCATCAATGAATCTCTCAA 152
|||||
792 leSerThrSerIleLeuAspLeuArgAspLysPheAsnGluSerLeuGln 808

153 GTGAATACTGCTGCTCATCCCAAGGAGCAACTCTGAGGAAGTCTT 202
|||||
809 ValAsnThrThrAlaLeuIleProLysGluAlaAsnSerGluGluValPh 825

203 TTTCTTTTAAACCAACAATTTACTTTTGAATAATGGCACAGATCTT 248
|||||
825 eLeuPheLysProGluAsnIleThrPheGluAsnGlyThrAspLeu 840

seq_name: sp_mammal:Q9TUB5

seq_documentation_block:
ID Q9TUB5 PRELIMINARY; PRT; 917 AA.
AC Q9TUB5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN.
GN AECG
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ILEAL MUCOSA;
RA Gaspar K.J., Gabriel S.E., Racette K.J., Forsyth G.W.;
RT "A cdna involved in porcine exocrine chloride conduction."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF095584; AAF00077.1; .
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.

SQ SEQUENCE 917 AA; 100735 MW; 649F41944F7EAD19 CRC64;

alignment_scores:
Quality: 348.00 Length: 82
Ratio: 4.640 Gaps: 0
Percent Similarity: 91.463 Percent Identity: 81.707

alignment_block:

US-09-049-696-14 x Q9TUB5 ..

Align seg 1/1 to: Q9TUB5 from: 1 to: 917

```

3 CTGAAGCGGGAATTCACGGGGGAGTCTCAATTAATCTGACTTGGACAGC 52
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
761 LeuLysAlaGlyIleGlnGlyAspAsnLeuValAsnLeuThrTrpThrAl 777

53 TCCTGGGGGATGATTATGACCATGGAACAGCTCACAAGTATATCATTCGAA 102
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
777 aProGlyAspAspTyrAspHisGlyArgAlaAspArgTyrIleIleArgI 794

103 TAAGTACAAGTATTCTTGATCTCAGAGACAAGTTCAATGAATCTCTTCAA 152
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
794 leSerThrAsnIleLeuAspLeuArgAspLysPheAsnAspSerValGln 810

153 GTGAATACTACTGCTCTCATCCCAAGGAGCCAACTCTGAGGAGTCTT 202
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
811 ValAsnThrThrAspLeuIleProLysGluAlaAsnSerGluGluValPh 827

203 TTTGTTTAAACACGAAACATTACTTTTGAATAATGGCACAGATCTT 248
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
827 evalPheGluLeuGlyIleProPheThrAsnGlyThrAspLeu 842

```

seq_name: sp_rodent:088826

```

seq_documentation_block:
ID Q88826 PRELIMINARY; PRT; 913 AA.
AC Q88826;
DT 01-NOV-1998 (TEMBLrel. 08, Created)
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE GOB-5 PROTEIN.
GN CLC3 OR GOB-5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RA Komiya T., Tanigawa Y., Hirohashi S.;
RT "Cloning of a gene, gob-5, which is expressed in intestinal goblet
cells in mice."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017156; BAA33743.1; -.
DR MGD; MGI:1346342; Clc3.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR02035; VFPA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VFPA; 1.
SQ SEQUENCE 913 AA; 100070 MW; A7FA2F9E1089806D CRC64;

```

alignment_scores:
Quality: 304.00 Length: 82
Ratio: 4.222 Gaps: 0
Percent Similarity: 87.805 Percent Identity: 71.951

alignment_block:

US-09-049-696-14 x O88826 ..

Align seg 1/1 to: O88826 from: 1 to: 913

```

3 CTGAAGCGGGAATTCACGGGGGAGTCTCAATTAATCTGACTTGGACAGC 52
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
759 LeuLysAlaSerIleGlnGlyGlnAsnLeuValAsnLeuThrTrpThrAl 775

53 TCCTGGGGGATGATTATGACCATGGAACAGCTCACAAGTATATCATTCGAA 102
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
775 aProGlyAspAspTyrAspHisGlyArgAlaSerAsnTyrIleIleArgm 792

103 TAAGTACAAGTATTCTTGATCTCAGAGACAAGTTCAATGAATCTCTTCAA 152
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
792 etSerThrSerIleValAspLeuArgAspHisPheAsnThrSerLeuGln 808

153 GTGAATACTACTGCTCTCATCCCAAGGAGCCAACTCTGAGGAGTCTT 202
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
809 ValAsnThrThrGlyLeuIleProLysGluAlaSerSerGluGluIlePh 825

203 TTTGTTTAAACACGAAACATTACTTTTGAATAATGGCACAGATCTT 248
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
825 eGluPheGluLeuGlyGlyAsnThrPheGlyAsnGlyThrAspIle 840

```

seq_name: sp_rodent:Q9D726

```

seq_documentation_block:
ID Q9D726 PRELIMINARY; PRT; 913 AA.
AC Q9D726;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE CHANNEL CALCIUM ACTIVATED 3.
GN CLC3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=STOMACH;
MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
PA Akawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Flieschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK008659; BAB25815.1; -.
DR MGD; MGI:1346342; Clc3.
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR02035; VFPA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VFPA; 1.
SQ SEQUENCE 913 AA; 100111 MW; 491E5854B06D9A89 CRC64;

```

alignment_scores:
Quality: 300.00 Length: 82
Ratio: 4.167 Gaps: 0
Percent Similarity: 87.805 Percent Identity: 70.732

alignment_block:

US-09-049-696-14 x Q9D726

Align seg 1/1 to: Q9D726 from: 1 to: 913

3 CTGAAGCGGGAATTCACGGGGCAGTCTCAATTAATCTGACTTGGACAGC 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
759 LeuAspAlaThrValHisGlnAsnLeuValAsnLeuThrTrpThrAl 775
53 TCCTGGGGATGATTATGACCATCGAACAGCTCACAAGTATATCATTCGAA 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
775 aProGlyAspGluTyAspHisGlyArgAlaSerAsnTyrIleIleArgM 792
103 TAAGTACAGTATTCTTGATCTCAGAGACAACTCAATGAATCTCTTCAA 152
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
792 etSerThrSerIleValAspLeuArgAspHisPheAsnThrSerLeuGln 808
153 GTGAATACTACTCTCTCATCCCAAGGAGCAACTCTGAGGAAGTCTT 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
809 ValAsnThrThrGlyLeuIleProLysGluAlaSerSerGluGluIlePh 825
203 TTTGTTTAAACCAAGAAATCTACTTTTGAATAATGCCACAGATCTT 248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
825 eGluPheGluLeuGlyGlyAsnThrPheGlyAsnGlyThrAspIle 840

seq_name: sp_human:Q9NXP1

seq_documentation_block:

ID Q9NXP1 PRELIMINARY; PRT: 469 AA.

AC Q9NXP1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE CDNA FLJ20131 FIS, CLONE COL06357.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Oka T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK000138; BAA90969.1;
SQ SEQUENCE 469 AA; 50694 MW; 6765986C886B3E56 CRC64;

alignment_scores:

Quality: 259.50 Length: 82
Ratio: 3.761 Gaps: 1
Percent Similarity: 84.146 Percent Identity: 63.415

alignment_block:

US-09-049-696-14 x Q9NXP1

Align seg 1/1 to: Q9NXP1 from: 1 to: 469

3 CTGAAGCGGGAATTCACGGGGCAGTCTCAATTAATCTGACTTGGACAGC 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
313 LeuAspAlaThrValHisGluAspIleIle...LeuThrTrpThrAl 328
53 TCCTGGGGATGATTATGACCATCGAACAGCTCACAAGTATATCATTCGAA 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
328 aProGlyAspAsnPheAspValGlyLysValGlnArgTyrIleIleArgI 345
103 TAAGTACAGTATTCTTGATCTCAGAGACAACTCAATGAATCTCTTCAA 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
345 leSerAlaSerIleLeuAspLeuArgAspSerPheAspAlaLeuGln 361

153 GTGAATACTACTCTCTCATCCCAAGGAGCAACTCTGAGGAAGTCTT 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
362 ValAsnThrThrAspLeuSerProLysGluAlaAsnSerLysGluSerPh 378
203 TTTGTTTAAACCAAGAAATCTACTTTTGAATAATGCCACAGATCTT 248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
378 eAlaPheLysProGluAsnIleSerGluGluAsnAlaThrHisIle 393

seq_name: sp_human:Q9UNF7

seq_documentation_block:

ID Q9UNF7 PRELIMINARY; PRT: 917 AA.

AC Q9UNF7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 2.
GN CAC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON;
RX MEDLINE=99364503; PubMed=10437792;
RA Agnel M., Vermet T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
RT chloride channel (CaCC) family predominantly expressed in the
RT digestive tract and trachea."
RL FEBS Lett. 455:295-301(1999).
DR EMBL: AF127035; AAD48398.1;
DR InterPro: IPR002035; VWFA.
DR Pfam: PF00092; vwa; 1.
DR PROSITE: PS50234; VWFA; 1.
DR SMART: SM00327; VWA; 1.
SQ SEQUENCE 917 AA; 101153 MW; EA01090D781BEB95 CRC64;

alignment_scores:

Quality: 259.50 Length: 82
Ratio: 3.761 Gaps: 1
Percent Similarity: 84.146 Percent Identity: 63.415

alignment_block:

US-09-049-696-14 x Q9UNF7

Align seg 1/1 to: Q9UNF7 from: 1 to: 917

3 CTGAAGCGGGAATTCACGGGGCAGTCTCAATTAATCTGACTTGGACAGC 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
761 LeuAspAlaThrValHisGluAspIleIle...LeuThrTrpThrAl 776
53 TCCTGGGGATGATTATGACCATCGAACAGCTCACAAGTATATCATTCGAA 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
776 aProGlyAspAsnPheAspValGlyLysValGlnArgTyrIleIleArgI 793
103 TAAGTACAGTATTCTTGATCTCAGAGACAACTCAATGAATCTCTTCAA 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
793 leSerAlaSerIleLeuAspLeuArgAspSerPheAspAlaLeuGln 809
153 GTGAATACTACTCTCTCATCCCAAGGAGCAACTCTGAGGAAGTCTT 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
810 ValAsnThrThrAspLeuSerProLysGluAlaAsnSerLysGluSerPh 826
203 TTTGTTTAAACCAAGAAATCTACTTTTGAATAATGCCACAGATCTT 248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
825 eAlaPheLysProGluAsnIleSerGluGluAsnAlaThrHisIle 841

seq_name: sp_rodent:O88860

seq_documentation_block:

ID O88860 PRELIMINARY; PRT: 901 AA.

```

O88860;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CHLORIDE CHANNEL CACC.
GN CLC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Romio L., Musante L., Cinti R., Moran O., Seri M., Galletta L.J.V.;
RT "Characterization of a murine gene homologous to the bovine CaCC
RT chloride channel."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF052746; AAC35003.1; -.
DR MGD; MGI:1316732; Clcal.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 901 AA; 100038 MW; 13DF068BF3337AAB CRC64;

alignment_scores:
  Quality: 210.50      Length: 82
  Ratio: 3.238        Gaps: 1
  Percent Similarity: 79.268      Percent Identity: 53.659

alignment_block:
US-09-049-696-14 x O88860 ..
  Align seg 1/1 to: O88860 from: 1 to: 901

3 CTGAAGCGGAAATTCACGGGGCAGTCTCATTATCTGACTTGGACAGC 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
764 LeuGluaLaGluPheIle...GlyAspTyrIleHisLeuThrTrpThral 779

53 TCCTGGGGATGATTATGACCATGGAACAGCTCACAAGTATATCATTCGAA 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
779 aProGlyLysValLeuAspAsnGlyArgAlaHisArgTyrIleIleArgm 796

103 TAAGTACAAGTATTCTTGATCTCAGACAGCAAGTTCAATGAATCTCTTCAA 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
796 etSerGlnHisProLeuAspLeuGlnGluAspPheAsnAlaThrLeu 812

153 GTGAATACTACTGCTCTCATCCCAAGGAGCCAACTCTGAGGAAGTCTT 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
813 ValAsnAlaSerSerLeuIleProLysGluAlaGlySerLysGluThrPh 829

203 TTTCTTTAAACCAGAAACATTTACTTTTGAATAATGGCAGACATCTT 248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
829 eLysPheLysProGluThrPheLysIleAlaAsnGlyIleGlnLeu 844

seq_name: sp_rodent:Q9QX15

seq_documentation_block:
ID Q9QX15 PRELIMINARY; PRT; 902 AA.
AC Q9QX15;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CALCIUM-SENSITIVE CHLORIDE CONDUCTANCE PROTEIN-1.
GN CLC1 OR MC1CAL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Gandhi R., Elble R.C., Gruber A.D., Schreier K.D., Ji H.-L.,
RX TISSUE-WHOLE LUNG; PubMed=9822685;
MEDLINE=99041980;

```

```

RA Fuller C.M., Pauli B.U.;
RT "Molecular and functional characterization of a calcium-sensitive
RT chloride channel from mouse lung."
RL J. Biol. Chem. 273:32096-32101(1998).
DR EMBL; AF047838; AAC79982.1; -.
DR MGD; MGI:1316732; Clcal.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 902 AA; 100107 MW; D9A915E698EA2D85 CRC64;

alignment_scores:
  Quality: 210.50      Length: 82
  Ratio: 3.238        Gaps: 1
  Percent Similarity: 79.268      Percent Identity: 53.659

alignment_block:
US-09-049-696-14 x Q9QX15 ..
  Align seg 1/1 to: Q9QX15 from: 1 to: 902

3 CTGAAGCGGAAATTCACGGGGCAGTCTCATTATCTGACTTGGACAGC 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
764 LeuGluaLaGluPheIle...GlyAspTyrIleHisLeuThrTrpThral 779

53 TCCTGGGGATGATTATGACCATGGAACAGCTCACAAGTATATCATTCGAA 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
779 aProGlyLysValLeuAspAsnGlyArgAlaHisArgTyrIleIleArgm 796

103 TAAGTACAAGTATTCTTGATCTCAGACAGCAAGTTCAATGAATCTCTTCAA 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
796 etSerGlnHisProLeuAspLeuGlnGluAspPheAsnAlaThrLeu 812

153 GTGAATACTACTGCTCTCATCCCAAGGAGCCAACTCTGAGGAAGTCTT 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
813 ValAsnAlaSerSerLeuIleProLysGluAlaGlySerLysGluThrPh 829

203 TTTCTTTAAACCAGAAACATTTACTTTTGAATAATGGCAGACATCTT 248
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
829 eLysPheLysProGluThrPheLysIleAlaAsnGlyIleGlnLeu 844

seq_name: sp_rodent:Q9QX15

seq_documentation_block:
ID Q9QX15 PRELIMINARY; PRT; 902 AA.
AC Q9QX15;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ENDOTHELIAL CHLORIDE CHANNEL.
GN CLCA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Elble R.C., Pauli B.;
RT "A murine endothelial homolog of Lu-ECAM-1 associated with lung
RT melanoma metastasis."
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF115852; AAG47626.1; -.
DR InterPro; IPR003961; FN_III.
DR SMART; SM00060; FN3; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 902 AA; 99855 MW; F2BA57DAA5D4AAAD CRC64;

alignment_scores:

```

Quality: 201.50 Length: 82
Ratio: 3.198 Gaps: 1
Percent Similarity: 76.829 Percent Identity: 52.439

alignment_block:
US-09-049-696-14 x Q9EQR4 ..
Align seg 1/1 to: Q9EQR4 from: 1 to: 902

3 CTGAGCGGGAATTCACGGGGGAGTCTCATTAATCTGACTTGGACAGC 52
|||||
764 LeuGluAlaGluPheIle...GlyAspTyrIleGlnLeuThrTrpThrAl 779
53 TCTGGGGATGATGATGACGATGGAACAGCTCACAAGTATATATCGAA 102
|||||
779 aProGlyLysValLeuAspLysGlyArgAlaHisArgTyrIleIleArgV 796
103 TAAGTACAAGTATCTTGATCTCAGACAGCAAGTTCAATGAATCTCTTCAA 152
|||||
796 alSerGlnHisProLeuGlyLeuGlnGluAspPheAsnAsnAlaThrLeu 812
153 GTGAATACTACTGCTCTCATCCCAAGGAGCAACTCTCGAGGAAGTCTT 202
|||||
813 ValAsnAlaSerSerLeuIleProLysGluAlaGlySerLysGluThrPh 829
203 TTGTTTAAACAGAAACATTAATCTTTGAAAATGGCAGAGATCTT 248
|||||
829 eLysPheLysProGluThrPheLysIleAlaAsnAspThrGlnLeu 844

seq_name: sp_rodent:Q9R070
seq_documentation_block:
ID Q9R070 PRELIMINARY; PRT: 902 AA.
AC Q9R070;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CA(2+)-SENSITIVE CHLORIDE CHANNEL 2.
GN CLCA2 OR CACC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY;
RX MEDLINE=20012773; PubMed=10544033;
RA Lee D., Ha S., Kho Y., Kim J., Cho K., Baik M., Choi Y.;
RT "Induction of mouse Ca(2+)-sensitive chloride channel 2 gene during
involution of mammary gland."; Commun. 264:933-937(1999).
RL Biochem. Biophys. Res. Commun. 264:933-937(1999).
DR EMBL: AF108501; AAF12731.1; -.
DR MGD: MGI:1931471; Clca2.
DR InterPro: IPR002035; VWFA.
DR SMART: SM00327; VWA: 1.
DR PROSITE: PS0234; VWFA: 1.
SQ SEQUENCE 902 AA; 99866 MW; 0FB746CF5B3D07F CRC64;

alignment_scores:
Quality: 198.50 Length: 82
Ratio: 3.202 Gaps: 1
Percent Similarity: 75.610 Percent Identity: 52.439

alignment_block:
US-09-049-696-14 x Q9R070 ..
Align seg 1/1 to: Q9R070 from: 1 to: 902

3 CTGAGCGGGAATTCACGGGGGAGTCTCATTAATCTGACTTGGACAGC 52
|||||
764 LeuGluAlaGluPheIle...GlyAspTyrIleGlnLeuThrTrpThrAl 779

53 TCCTGGGATGATGATGACCATGGAACAGCTCACAAGTATATATCATTCGAA 102
|||||
779 aProGlyLysValLeuAspLysGlyArgAlaHisArgTyrIleIleArgV 796
103 TAAGTACAAGTATCTTGATCTCAGACAGCAAGTTCAATGAATCTCTTCAA 152
|||||
796 alSerGlnHisProLeuGlyLeuGlnGluAspPheAsnAsnAlaThrLeu 812
153 GTGAATACTACTGCTCTCATCCCAAGGAGCAACTCTCGAGGAAGTCTT 202
|||||
813 ValAsnAlaSerSerLeuIleProLysGluAlaGlySerLysGluThrPh 829
203 TTGTTTAAACAGAAACATTAATCTTTGAAAATGGCAGAGATCTT 248
|||||
829 eLysPheLysProGluThrPheLysIleAlaAsnAspThrGlnLeu 844

seq_name: sp_mammal:O18741
seq_documentation_block:
ID O18741 PRELIMINARY; PRT: 905 AA.
AC O18741;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elble R.C., Widom J., Gruber A.D., Abdel-Chany M., Levine R.,
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF001261; AAB86529.1; -.
DR InterPro: IPR002035; VWFA.
DR InterPro: IPR003961; FN_III.
DR Pfam: PF00041; fn3; 1.
DR PROSITE: PS0234; VWFA: 1.
DR SMART: SM00327; VWA: 1.
SQ SEQUENCE 905 AA; 101005 MW; D4CA6E7919115E88 CRC64;

alignment_scores:
Quality: 195.00 Length: 70
Ratio: 3.482 Gaps: 0
Percent Similarity: 80.000 Percent Identity: 55.714

alignment_block:
US-09-049-696-14 x O18741 ..
Align seg 1/1 to: O18741 from: 1 to: 905
33 ATTAATCTGACTTGGACAGCTCTGGGGATGATGATGACCATCGAACACC 82
|||||
777 IleGlnLeuSerTrpThrAlaProGlyAsnValLeuAspLysGlyLysAl 793
83 TCACAAGTATATCATTCGAATACAAAGTATCTTGTATCTCAGACAGA 132
|||||
793 aAsnSerTyrIleIleArgIleSerLysSerPheMetAspArgGlnGlu 810
133 AGTTCAATGAATCTCTTCAAGTGAATACTACTGCTCTCATGCCAAGGAA 182
|||||
810 spPheAspAsnAlaThrLeuValAsnThrSerAsnLeuIleProLysGlu 826
183 GCCAACTCTGAGGAAGTCTTTTGTAAACAGCAAAAACATTACTTTTGA 232
|||||
827 AlaGlySerLysGluAsnPheGluPheLysProGluHisPheArgValGI 843
233 AAATGGCACA 242
|||||

This Page Blank (uspto)


```

RT truncated protein isoform in vivo."
RL Mol. Genet. Metab. 70:69-74(2000).
CC -1- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER
CC MEMBRANE THAT ALLOWS DIFFUSION OF SMALL HYDROPHILIC MOLECULES
CC (BY SIMILARITY)
CC -1- SUBCELLULAR LOCATION: OUTER MITOCHONDRIAL MEMBRANE.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED, HIGHEST IN TESTIS.
CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF038962; AAC39876.1;
DR EMBL; U90943; AAB93872.1;
DR EMBL; AF151682; AAD49610.1;
DR EMBL; AF151679; AAD49610.1; JOINED.
DR EMBL; AF151681; AAD49610.1; JOINED.
DR InterPro: IPR001925; Euk_porin.
DR Pfam: PF01459; Euk_porin; 1.
DR PRINTS; PR00185; EUKARYOTIC_PORIN.
DR PROSITE; PS00558; EUKARYOTIC_PORIN; 1.
DR Outer membrane; Porin; Mitochondrion; Alternative splicing.
KW VARSPLIC 39 V -> VM (IN ISOFORM 2).
FT
SQ SEQUENCE 283 AA; 30658 MW; E03CBCEDA72A9783 CRC64;

alignment_scores:
  Quality: 69.00 Length: 62
  Ratio: 2.156 Gaps: 2
Percent Similarity: 51.613 Percent Identity: 32.258

alignment_block:
US-09-049-696-14 x POR3_HUMAN
Align seg 1/1 to: POR3_HUMAN from: 1 to: 283

21 GGGGGCAGTCTC.....ATTAA 37
|||||
191 GlyGlySerIleGlnLysValAsnGluLysIleGluThrSerIleAs 207
38 TGTGACTTGGACAGCTCCTGGGATGATTGACCATGGAACAGCTCACA 87
|||||
207 nLeuAlaIleThrAlaGlySerAsnAsnThrArgPheGlyIleAlaIle 224
88 AGTATATCATTCGAATAAGTACAAGTATTCTTGTCTGATCTCAGACACAAGTTC 137
|||||
224 ystyrMetLeuaspCysArgThrSerLeu..... 233

138 AATGAATCTCTTCAAGTGAATCTACTGCTCTCATC 173
|||
234 .....SerAlaLysValAsnAsnAlaSerLeuIle 243

seq_name: SwissProt_39:POLG_TMEVD
seq_documentation_block:
ID POLG_TMEVD STANDARD; PRT; 2301 AA.
AC P13899; Q88564; Q88565; Q88567; Q88568; Q88569; Q88570;
AC Q88571; Q88572; Q88573; Q88574; Q88580;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; P3A; GENOME-LINKED PROTEIN VP3; PICORNAIN 3C
DE (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
DE (EC 2.7.7.48)].

```

```

RT Ratio: 3.333 Gaps: 0
Percent Similarity: 80.282 Percent Identity: 54.930
alignment_block:
US-09-049-696-14 x ECLC_BOVIN
Align seg 1/1 to: ECLC_BOVIN from: 1 to: 903

33 ATTAATCTGACTTGGACAGCTCTGGGGATGATTGACCATGGACACACC 82
|||||
775 IleGlnLeuSerTrpThrAlaProAlaAsnValLeuAspLysGlyLysAl 791

83 TCACAGTATATCATCGAATAGTACAAGTATCTTGATCTCAGAGACA 132
|||||
791 aAsnSerTyrllelleGargileSerLysSerPheLeuAspLeuGlnLysA 808

133 AGTTCAATGAATCTCTCAAGTGAATCAATGCTCTCATCCCAAGGAA 182
|||||
808 spPheAspAsnAlaThrLeuValAsnThrSerSerLeuLysProLysGlu 824

183 GCCAACTCTCAGGAAGTCTTTTGTGTTTAAACACAGAAACATTAATTGCA 232
|||||
825 AlaGlySerAspGluAsnPheGluPheLysProGluPropheargileG 841

233 AAATGGCACAGAT 245
|||||
841 uAsnGlyThrAsn 845

seq_name: SwissProt_39:POR3_HUMAN
seq_documentation_block:
ID POR3_HUMAN STANDARD; PRT; 283 AA.
AC Q9Y277; Q9UIS0;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN 3 (VDAC-3) (HVDAC3)
DE [OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN 3].
GN VDAC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=98318631; PubMed=9653160;
RA Mao M., Fu G., Wu J.-S., Zhang Q.-H., Zhou J., Kan L.-X., Huang Q.-H.,
RA He K.-L., Gu B.-W., Han Z.-G., Shen Y., Gu J., Yu Y.-P., Xu S.-H.,
RA Wang Y.-X., Chen S.-J., Chen Z.;
RA "Identification of genes expressed in human CD34(+) hematopoietic
RA stem/progenitor cells by expressed sequence tags and efficient full-
RA length cDNA cloning."
RL Proc. Natl. Acad. Sci. U.S.A. 95:8175-8180(1998).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=98434325; PubMed=9781040;
RA Rahmani Z., Maunoury C., Siddiqui A.;
RA "Isolation of a novel human voltage-dependent anion channel gene."
RL Eur. J. Hum. Genet. 6:337-340(1998).
RN [3]
RN SEQUENCE OF 1-253 FROM N.A.
RX MEDLINE=99431679; PubMed=10501981;
RA Decker W.K., Bowles K.R., Schatte E.C., Towbin J.A., Craigen W.J.;
RA "Revised fine mapping of the human voltage-dependent anion channel
RA loci by radiation hybrid analysis."
RL Mamm. Genome 10:1041-1042(1999).
RN [4]
RN ALTERNATIVE SPLICING.
RX MEDLINE=20295349; PubMed=10833333;
RA Decker W.K., Craigen W.J.;
RA "The tissue-specific, alternatively spliced single ATG exon of the
RA type 3 voltage-dependent anion channel gene does not create a

```


OS Theller's murine encephalomyelitis virus (strain DA) (TMEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Cardiovirus.
OX NCBI_TaxID=12126;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88206072; PubMed=2834872;
RA Ohara Y., Stein S., Fu J., Stillman L., Klamon L., Roos R.P.;
RT "Molecular cloning and sequence determination of DA strain of
RT Theller's murine encephalomyelitis viruses.";
RL Virology 164:245-255(1988).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=92196057; PubMed=1549565;
RA Grant R.A., Filman D.J., Fujinami R.S., Icenogle J.P., Hogle J.M.;
RT "Three-dimensional structure of Theller virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:2061-2065(1992).
CC -!- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR
CC MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: CLOSELY RELATED TO ENCEPHALOMYOCARDITIS VIRUS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M20301; AAA47928.1; -
DR PIR; A31228; GNNYTN.
DR PDB; ITME; 3I-JAN-94.
DR MEROPS; C03.009; -
DR MEROPS; U29.001; -
DR InterPro; IPR000605; RNA_helicase.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR001676; Rhv.
DR Pfam; PF00073; rhv; 3.
DR Pfam; PF00860; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
KW Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate;
KW 3d-structure.
FT PROPEP 1 76 LEADER PEPTIDE.
FT CHAIN 77 147 COAT PROTEIN VP4 (P1A).
FT CHAIN 148 414 COAT PROTEIN VP2 (P1B).
FT CHAIN 415 646 COAT PROTEIN VP3 (P1C).
FT CHAIN 647 920 COAT PROTEIN VP1 (P1D).
FT CHAIN 921 1062 CORE PROTEIN P2A.
FT CHAIN 1063 1189 CORE PROTEIN P2B.
FT CHAIN 1190 1515 CORE PROTEIN P2C.
FT CHAIN 1516 1603 CORE PROTEIN P3A.
FT CHAIN 1604 1623 GENOME-LINKED PROTEIN VPG (P3B).
FT CHAIN 1624 1840 PICORNAIN 3C.
FT CHAIN 1841 2301 RNA-DIRECTED RNA POLYMERASE P3D.
FT LIPID 77 77 MYRISTATE (BY SIMILARITY).
FT ACT_SITE 1791 1791 PROTEASE (POTENTIAL).
FT ACT_SITE 1809 1809 PROTEASE (POTENTIAL).
FT TURN 96 97
FT HELIX 103 106
FT STRAND 109 110
FT STRAND 162 166
FT TURN 167 168
FT STRAND 169 173
FT STRAND 179 181
FT HELIX 182 184

193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

	alignment_scores:			
DR	Quality:	67.00	Length:	78
KW	Ratio:	1.457	Gaps:	4
	Percent Similarity:	58.974	Percent Identity:	26.923
	alignment_block:			
US-09-049-896-14 x SR72_HUMAN				
	Align seg 1/1 to: SR72_HUMAN from: 1 to: 670			
	12 GAATTCACGGGCGAGCTCATTAACTCGACTTGCACAGCTCCTGGGGA	61		
	::: ::::	:	:	:
	112 GlutLeuryrGlycInvalleutyrrArgIeu.....IGL	122		
	62 TGATTGTGCCCATGGCAACAGCT...CACAAAGTATATCATTTCCGAATAAAGA	108		

DR	InterPro: IPR001925; Euk_porin.
DR	Pfam: PF01459; Euk_porin; 1.
DR	PRINTS: PR00185; EUKARYTPORIN.
DR	PROSITE: PS00558; EUKARYOTIC_PORIN; 1.
DR	Outer membrane; porin; Mitochondrion.
DR	SEQUENCE 283 AA; 30739 MW; D3U5DA2EE42BEC3B CRC64;
DR	SW
DR	KW

alignment_scores:		
Quality:	65.00	Length: 62
Ratio:	2.097	Gaps: 2
Percent Similarity:	50.000	Percent Identity: 32.258

alignment_block:

Align seg 1/1 .to: POR3_BOVIN from: 1 to: 283

21. GGGGGCAGTC.....ATTAA 37
|||||:::
|||c|c|g|s|o|f|l|e|t|r|y|c|l|i|n|v|s|v|a|l|a|s|n|g|l|i|n|v|s|l|e|g|l|u|t|h|r|s|e|r|i|e|a|s 207

191 GlyGlySerIleTyrGlnLysValAsnGluLysIleGluThrSerIleAs 20
38 TCTGACTTGGACAGCTCCTGGGGATGATTTATGACCATTGGAACAGCTCACA 87

207 nLeuAlaTrpThrAlaGlySerAsnAsnThrArgPheGlyIleAlaAlaL 224

88 AGTATATCATTCGAATAAGTACAAGTATTCTTGATCTCAGAGACAAGTTC 137
||||| :::: |||||:::
224 vettur1.vst1.ou.aa00.cvs.aa00.thrSer1.ell 233

224 ystYrLysLeuAspCysArgThrSerLeu..... 233
139 AATGAATCTCTTCAAGTGAATACTACTGCTCTCATC 173

```

234 .....SerAlaLysValAsnAsnAlaSerLeuIle 243
      |||  ::||| ||| ::| ::| ::| ::| ::| ::|

```

seq|name:|SwissProt|39:POR3_MOUSE

```
seq_documentation_block:
ID:  PCB3 MOUSE          STANDARD:
PRT: 283 AA.
```

Q60931; AC
15-JUL-1999 (Rel. 38, Created)
15-JUL-1999 (Rel. 38, Last sequence update)
20-AUG-2001 (Rel. 40, Last annotation update)
VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN 3 (VDAC-3) (MVDAC3)
COUPLER, MITOCHONDRIAL MEMBRANE PROTEIN PORIN 3.

DT | 20-AUG-2001 (Rel. 40, Last annotation update)
DE | VOLTAGE-DEPENDENT ANTON-SELECTIVE CHANNEL PROTEIN 3 (VDAC-3) (MVDAC3)
PE | CITTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN 3).

OS Mus musculus (Mouse).
OS Chondrichthyes: Vorticeata: Euteleostomi:
OS Osteichthyes: Osteichthyes: Vorticeata: Euteleostomi:

[illegible]

RN	[1]	SEQUENCE FROM N.A.
RP		TISSUE=Liver:
RC		

MEJLINE-96411667; PubMed=8812436;
Sampson M.J., Lovell R.S., Davison D.B., Craig W.J.;
Sampson M.J., Lovell R.S., Davison D.B., Craig W.J.;

RT "A novel mouse mitochondrial voltage-dependent anion channel gene localizes to chromosome 8." RT localizes to chromosome 8." RT Genomics 36:192-196(1996). RL

CC FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER
CC MEMBRANE THAT ALLOWS DIFFUSION OF SMALL HYDROPHILIC MOLECULES
CC (BY SIMILARITY).

CC	-	SUBCELLULAR LOCATION: OUTER MITOCHONDRIAL MEMBRANE.
CC	-	SUBCELLULAR LOCATION: OUTER MITOCHONDRIAL MEMBRANE.
CC	-	TISSUE SPECIFICITY: HIGHEST LEVELS OF EXPRESSION DETECTED IN LIVER, KIDNEY, PANCREAS, AND TESTIS.
CC	-	TISSUE SPECIFICITY: HIGHEST LEVELS OF EXPRESSION DETECTED IN LIVER, KIDNEY, PANCREAS, AND TESTIS.
CC	-	TISSUE SPECIFICITY: HIGHEST LEVELS OF EXPRESSION DETECTED IN LIVER, KIDNEY, PANCREAS, AND TESTIS.

CC TESTS, LESS BUT STILL ABUNDANT EXPRESSION IN HEART, KIDNEY, LIVER, PANCREAS, SPLEEN, STOMACH, SMALL INTESTINE, COLON, UTERUS, BLADDER, AND SKELTAL MUSCLE.

CC -!- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEET.

CC
CC
CC
CC
CC

-- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY
THIS SWISS-PROT ENTRY IS COPYRIGHTED. IT IS PRODUCED THROUGH A COLLABORATION WITH THE EMBL DATA BANK.

between the Swiss Institute of Bioinformatics and the EMBL out-
the European Bioinformatics Institute. There are no restrictions on
its content is in the public domain and can be used by any research
specific institutions as long as its content is in the public domain and
can be used by any research institution.

CC BY-NC-SA 4.0 International license. All rights reserved. Use of this document is restricted to non-profit institutions as long as the content is not modified and this statement is not removed. Usage by and for other entities requires a license agreement (See <http://www.isb-sib.ch/>)

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U30839; AAB47776.1; -.
DR MGD; MGI:106922; Vdac3.
DR InterPro; IPR001925; Euk_porin.
DR Pfam; PF01459; Euk_porin; 1.
DR PRINTS; PR00185; EUKARYTOPORIN.
DR PROSITE; PS00558; EUKARYOTIC_PORIN; 1.
DR Outer membrane; Porin; Mitochondrion.
KW Outer membrane; Porin; Mitochondrion.
SQ SEQUENCE 283 AA; 30753 MW; A95AFD67C611228C CRC64;

alignment_scores:
  Quality: 65.00 Length: 62
  Ratio: 2.097 Gaps: 2
Percent Similarity: 50.000 Percent Identity: 32.258

alignment_block:
US-09-049-696-14 x POR3_MOUSE ..
Align seg 1/1 to: POR3_MOUSE from: 1 to: 283
21 GGGGGCAGTCTC.....ATTAA 37
   |||||
191 GlyGlySerIleTyrGlnLysValAsnGluArgIleGluThrSerIleAs 207
38 TCTGACTGGACAGCTCTCTGGGATGATTATGACCATGGAACAGCTCACA 87
   |||||
207 nLeuAlaTrpThrAlaGlySerAsnAsnThrArgPheGlyIleAlaAla 224
88 AGTATATCATTCGAAAGTACAAAGTATTCTTGATCTCAGACAGCAAGTTC 137
   |||||
224 ystYrLysLeuAspCysArgThrSerLeu..... 233
138 AATGAATCTCTCAAGTAACTACTGCTCTCATC 173
   |||||
234 .....SerAlaLysValAsnAsnAlaSerLeulle 243

seq_name: SwissProt_39:POR3_MOUSE

seq_documentation_block:
ID POR3_MOUSE STANDARD; PRT; 283 AA.
AC Q9R120; Q9WTU2; Q9ESR2; Q9J131;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN 3 (VDAC-3) (OUTER
DE MITOCHONDRIAL MEMBRANE PROTEIN PORIN 3).
GN VDAC3.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW ZEALAND WHITE; TISSUE=Cornea;
RA Rae J.L.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -!- DOMAIN: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER
CC -!- FUNCTION: THAT ALLOWS DIFFUSION OF SMALL HYDROPHILIC MOLECULES
CC (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: OUTER MITOCHONDRIAL MEMBRANE.
CC -!- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF209727; AAF22837.1; -.
DR InterPro; IPR001925; Euk_porin.
DR Pfam; PF01459; Euk_porin; 1.
DR PRINTS; PR00185; EUKARYTOPORIN.
DR PROSITE; PS00558; EUKARYOTIC_PORIN; 1.
DR Outer membrane; Porin; Mitochondrion.
KW Outer membrane; Porin; Mitochondrion.
SQ SEQUENCE 283 AA; 30651 MW; BC0C5616366090A0 CRC64;

alignment_scores:
  Quality: 65.00 Length: 62
  Ratio: 2.097 Gaps: 2
Percent Similarity: 50.000 Percent Identity: 32.258

alignment_block:
US-09-049-696-14 x POR3_RABBIT ..
Align seg 1/1 to: POR3_RABBIT from: 1 to: 283
21 GGGGGCAGTCTC.....ATTAA 37
   |||||
191 GlyGlySerIleTyrGlnLysValAsnGluLysIleGluThrSerIleAs 207
38 TCTGACTGGACAGCTCTCTGGGATGATTATGACCATGGAACAGCTCACA 87
   |||||
207 nLeuAlaTrpThrAlaGlySerAsnAsnThrArgPheGlyIleAlaAla 224
88 AGTATATCATTCGAAAGTACAAAGTATTCTTGATCTCAGACAGCAAGTTC 137
   |||||
224 ystYrLysLeuAspCysArgThrSerLeu..... 233
138 AATGAATCTCTCAAGTAACTACTGCTCTCATC 173
   |||||
234 .....SerAlaLysValAsnAsnAlaSerLeulle 243

seq_name: SwissProt_39:POR3_RABBIT

seq_documentation_block:
ID POR3_RABBIT STANDARD; PRT; 283 AA.
AC Q9R120; Q9WTU2; Q9ESR2; Q9J131;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL PROTEIN 3 (VDAC-3) (RVDAC3)
DE (OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN 3).
GN VDAC3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC MEDLINE=20453129; PubMed=10998068;
RA Shinohara Y., Ishida T., Hino M., Yamazaki N., Baba Y., Terada H.;
RT 'Characterization of porin isoforms expressed in tumor cells.';
RL Eur. J. Biochem. 267:6067-6073(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=Sprague-Dawley;
RA Rae J.L.;
RT 'Ion channels in the lens.';
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 10-283 FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Heart;
RX MEDLINE=98390268; PubMed=9714728;
RA Anflous K., Blondel O., Bernard A., Khrestchatsky M.,
RA Ventura-Clapier R.;
RT 'Characterization of rat porin isoforms: cloning of a cardiac type-3
RT variant encoding an additional methionine at its putative N-terminal
RT region.';
RL Biochim. Biophys. Acta 1399:47-50(1998).
CC -!- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER
```

MEMBRANE THAT ALLOWS DIFFUSION OF SMALL HYDROPHILIC MOLECULES
(BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: OUTER MITOCHONDRIAL MEMBRANE.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1/RVDAC3 (SHOWN HERE) AND
CC 2/RVDAC3V; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: RVDAC3 IS WIDELY EXPRESSED, STRONGEST IN
CC ATRIA. RVDAC3V IS ALSO WIDELY EXPRESSED, HIGHEST IN BRAIN BUT NOT
CC EXPRESSED IN KIDNEY.
CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

EMBL; AB039664; BAB13475.1; -
DR EMBL; AF268469; AAF80117.1; -
DR EMBL; AF048829; AAD2722.1; -
DR EMBL; AF048830; AAD2723.1; -
DR InterPro; IPR001925; Euk_porin.
DR Pfam; PF01459; Euk_porin; 1.
DR PROSITE; PS00558; EUKARYOTIC_PORIN; 1.
KW Outer membrane; Porin; Mitochondrion; Alternative splicing.
FT VARSPIC 39 39 V -> VM (IN ISOFORM 2).
FT CONFLICT 128 128 K -> N (IN REF. 2).
SQ SEQUENCE 283 AA; 30798 MW; 3800246686557864 CRC64;

alignment_scores:
Quality: 65.00 Length: 62
Ratio: 2.097 Gaps: 2
Percent Similarity: 50.000 Percent Identity: 32.258

alignment_block:

US-09-049-696-14 x POR3_RAT ..
Align seg 1/1 to: POR3_RAT from: 1 to: 283
21 GGGGGCAGCTCTCATTAA 37
|||||
191 GlyGlySerIleTyGlnArgValAsnGluLysIleGluThrSerIleAs 207
38 TCTGACTGGACAGCTCTCGGGGATGATTATGACCATGGAACAGCTCACA 87
|||||
207 nLeuAlaTrpThrAlaGlySerAsnThrArgPheGlyIleAlaAla 224
88 AGTATATCATTCGATTAAGTACAGATTCTTGATCTCAGAGACAGTTC 137
|||||
224 ystYrArgLeuAspCysArgThrSerLeu 233
138 AATGAATCTCTTCAAGTCAATACACTACTGCTCTCATC 173
|||||
234SerAlaLysValAsnAsnAlaSerLeulle 243

seq_name: SwissProt_39:MGR8_HUMAN

seq_documentation_block:

ID MGR8_HUMAN STANDARD; PRT; 908 AA.
AC Q00222; O15493; O95945; O95946;
DT 01-NOV-1997 (Rel. 35, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 8 PRECURSOR.
GN GRM8 OR GPCR1H OR MGLUR8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

SEQUENCE FROM N.A. (ISOFORM A).
MEDLINE-98141892; PubMed-9473604;
RA Wu S., Wright R.A., Rockey P.K., Burchett S.G., Arnold J.S.,
RA Rebeck P.R. Jr., Johnson B.G., Schoepf D.D., Belagaje R.M.;
RA Group III human metabotropic glutamate receptors 4, 7 and 8:
RT molecular cloning, functional expression, and comparison of
RT pharmacological properties in RGT cells.;
RL Brain Res. Mol. Brain Res. 53:88-97(1998).
[2]
RN SEQUENCE FROM N.A. (ISOFORM A).
RP MEDLINE-97446143; PubMed-9299241;
RX Scherer S.W., Soder S., Duvoisin R.M., Huizenga J.J., Tsui L.-C.;
RA "The human metabotropic glutamate receptor 8 (GRM8) gene: a
RT disproportionately large gene located at 7q31.3-q32.1.";
RL Genomics 44:232-236(1997).
[3]
RN SEQUENCE FROM N.A. (ISOFORMS B AND C).
RP TISSUE-Fetal brain;
RX Melherbe P., Kratzelisen C., Lundstrom K., Richards J.G., Faull R.L.M.,
RA "Cloning and functional expression of alternative spliced variants of
RT the human metabotropic glutamate receptor 8";
FL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR
CC IS MEDIATED BY A G-PROTEIN THAT INHIBITS ADENYLATE CYCLASE
CC ACTIVITY.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A (SHOWN HERE), B AND C; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

EMBL; U92459; AAB51764.1; -
DR EMBL; U95025; AAB72040.1; -
DR EMBL; AJ236921; CAB36968.1; -
DR EMBL; AJ236922; CAB36969.1; -
DR GCRdb; GCR_1889; -
DR GCRdb; GCR_2604; -
DR MIM; 601116; -
DR InterPro; IPR001828; ANF_receptor.
DR InterPro; IPR000337; GPCR_Mgr.
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCR_MGR.
DR PRINTS; PR00593; MTABOTROPICR.
DR PRINTS; PR01058; MTABOTROPICR.
DR PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.
DR PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.
DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
DR PROSITE; PS0259; G_PROTEIN_RECEP_F3_4; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Multigene family; Olfaction; Alternative splicing.
FT SIGNAL 1 33
FT CHAIN 34 908
FT DOMAIN 34 583
FT TRANSMEM 584 608
FT DOMAIN 609 620
FT TRANSMEM 621 641
FT DOMAIN 642 647
FT TRANSMEM 648 668
FT DOMAIN 669 695
FT TRANSMEM 696 716
FT DOMAIN 717 746
FT TRANSMEM 747 768
FT DOMAIN 769 781
FT TRANSMEM 782 803

POTENTIAL.
METABOTROPIC GLUTAMATE RECEPTOR 8.
EXTRACELLULAR (POTENTIAL).
I (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
II (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
III (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
IV (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
V (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
VI (POTENTIAL).

```
FT DOMAIN 804 818 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 819 843 VII (POTENTIAL).
FT DOMAIN 844 908 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 452 452 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 480 480 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 565 565 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 893 908 TSSTKTYISYSHSI -> SKSSVEFPVMSGSTS (IN
FT ISOFORM B).
FT VARSPLIC 454 501 SAGTPVTFNENGDPAGRYDIFQIITNKSTYKVIQHWINO
FT LHLKVED -> CRRGIOMSLPWPPTLTPFSFSSMAVLILS
FT LMKTEMLLDVMISSIK (IN ISOFORM C).
FT MISSING (IN ISOFORM C).
FT VARSPLIC 502 908 R -> A (IN REF. 2).
FT CONFLICT 194 194 A -> I (IN REF. 2).
FT CONFLICT 460 460 A -> G (IN REF. 2).
FT CONFLICT 642 642 A -> N (IN REF. 1).
FT CONFLICT 768 768 S -> T (IN REF. 2).
FT CONFLICT 904 904 S -> T (IN REF. 2).
SQ SEQUENCE 908 AA; 101740 MW; 95C2D5883DAF6DE CRC64;

alignment_scores:
  Quality: 65.00 Length: 44
  Ratio: 2.167 Gaps: 1
Percent Similarity: 68.182 Percent Identity: 29.545

alignment_block:
US-09-049-696-14/rev x MGR8_HUMAN ..

Align seg 1/1 to: MGR8_HUMAN from: 1 to: 908

131 GTCTCTGAGATCAAGATACCTGTACT.....TATTCGAATGATATACT 88
:|||||:|||||:|||||:|||||:
744 IleSerAspLeuSerLeuIleCysSerLeuGlyTyrSerIleLeuLeuMe 760
87 TGTGAGCTGTTCCTGGTCATCAATCAATCCCGAGGAGCTGCCAAGTCAGA 38
:|||||:|||||:|||||:|||||:
760 tValThrCysThrValTyrAlaIleLysThrArgGlyValProGluThrp 777
37 TTAATGAGACGCCCCCGCTGAATTCGCCCTT 6
:|||||:|||||:|||||:|||||:
777 heAsnGluAlaLysProIleGlyPheThrMet 787

seq_name: SwissProt_39:MGR8_MOUSE

seq_documentation_block:
ID MGR8_MOUSE STANDARD; PRT; 908 AA.
AC P47743;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 8 PRECURSOR.
GN GRM8 OR GPRC1H OR MGLUR8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Retina;
RX MEDLINE=95239344; PubMed=7722646;
RA Duvoisin R.M., Zhang C., Ramonell K.;
RT "A novel metabotropic glutamate receptor expressed in the retina and
RT olfactory bulb.;"
RL J. Neurosci. 15:3075-3083(1995).
CC -!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR
CC IS MEDIATED BY A G-PROTEIN THAT INHIBITS ADENYLATE CYCLASE
CC ACTIVITY.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN OLFACTORY BULB.
CC ACCESSORY OLFACTORY BULB, AND MAMMILLARY BODY. WEAKER EXPRESSION
CC IN THE RETINA, AND IN SCATTERED CELLS IN THE CORTEX AND HINDBRAIN.
```

```
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensed@isb-sib.ch).
CC -----
CC EMBL: U17252; AAA68149.1; .
CC GCRDB: GCR_1712; .
CC MGI: 1351345; Gprclh.
CC InterPro: IPR001828; ANF_receptor.
CC InterPro: IPR000337; GPCR_Mgr.
CC Pfam: PF00003; 7tm_3; 1.
CC PRINTS: PF01094; ANF_receptor; 1.
CC PRINTS: PR00248; GPCRMR.
CC PRINTS: PR00593; MTABOTROPICR.
CC PRINTS: PR01058; MTABOTROPICR.
CC PROSITE: PS00979; G_PROTEIN_RECP_F3_1; 1.
CC PROSITE: PS00980; G_PROTEIN_RECP_F3_2; 1.
CC PROSITE: PS00981; G_PROTEIN_RECP_F3_3; 1.
CC PROSITE: PS0259; G_PROTEIN_RECP_F3_4; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Multigene family; Olfaction.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 908 METABOTROPIC GLUTAMATE RECEPTOR 8.
FT LC/AIN 34 583 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 584 608 I (POTENTIAL).
FT DOMAIN 609 620 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 621 641 II (POTENTIAL).
FT DOMAIN 642 647 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 648 668 III (POTENTIAL).
FT DOMAIN 669 695 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 696 716 IV (POTENTIAL).
FT DOMAIN 717 746 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 747 768 V (POTENTIAL).
FT DOMAIN 769 781 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 782 803 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 804 818 VI (POTENTIAL).
FT TRANSMEM 819 843 VII (POTENTIAL).
FT DOMAIN 844 908 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 452 452 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 480 480 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 565 565 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 908 AA; 101413 MW; 4CDD9D35827ED41F CRC64;

alignment_scores:
  Quality: 65.00 Length: 44
  Ratio: 2.167 Gaps: 1
Percent Similarity: 68.182 Percent Identity: 29.545

alignment_block:
US-09-049-696-14/rev x MGR8_MOUSE ..

Align seg 1/1 to: MGR8_MOUSE from: 1 to: 908

131 GTCTCTGAGATCAAGATACCTGTACT.....TATTCGAATGATATACT 88
:|||||:|||||:|||||:|||||:
744 IleSerAspLeuSerLeuIleCysSerLeuGlyTyrSerIleLeuLeuMe 760
87 TGTGAGCTGTTCCTGGTCATCAATCAATCCCGAGGAGCTGCCAAGTCAGA 38
:|||||:|||||:|||||:|||||:
760 tValThrCysThrValTyrAlaIleLysThrArgGlyValProGluThrp 777
37 TTAATGAGACGCCCCCGCTGAATTCGCCCTT 6
:|||||:|||||:|||||:|||||:
777 heAsnGluAlaLysProIleGlyPheThrMet 787
```

```
seq_name: SwissProt_39:MGR8_RAT
seq_documentation_block:
ID MGR8_RAT STANDARD; PRT; 908 AA.
AC P70579;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE METABOTROPIC GLUTAMATE RECEPTOR 8 PRECURSOR.
GN GRM8 OR GPRC1H OR MGLUR8.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97168760; PubMed=9016353;
RA Saugstad J.A., Kinzie J.M., Shinohara M.M., Segerson T.P.,
RA Westbrook G.L.;
RT "Cloning and expression of rat metabotropic glutamate receptor 8
RT reveals a distinct pharmacological profile."
RL Mol. Pharmacol. 51:119-125(1997).
CC -!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR
CC IS MEDIATED BY A G-PROTEIN THAT INHIBITS ADENYLATE CYCLASE
CC ACTIVITY.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: PROMINENT EXPRESSION IN OLFACTORY BULB,
CC PONTINE GRAY, LATERAL RETICULAR NUCLEUS OF THE THALAMUS, AND
CC PIRIFORM CORTEX. LESS ABUNDANT EXPRESSION IN CEREBELLAR CORTEX,
CC HIPPOCAMPUS, CEREBELLUM, AND MAMMARY BODY.
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U63288; AAC09537.1; .
CC GCRDB; GCR_1411; .
CC InterPro: IPR001828; ANF_receptor.
CC InterPro: IPR000337; GPCR_Mgr.
CC Pfam: PF00003; 7tm3; 1.
CC PRINTS; PR00248; GPCRMR.
CC PRINTS; PR00593; MTABOTROPICR.
CC PRINTS; PR01058; MTABOTROPICR.
CC PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.
CC PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.
CC PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
CC PROSITE; PS00982; G_PROTEIN_RECEP_F3_4; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
CC Multigene family; Olfaction.
CC SIGNAL 1 33
CC CHAIN 34 908
CC DOMAIN 34 583
CC TRANSMEM 584 608
CC DOMAIN 609 620
CC TRANSMEM 621 641
CC DOMAIN 642 647
CC TRANSMEM 648 668
CC DOMAIN 669 695
CC TRANSMEM 696 716
CC DOMAIN 717 746
CC TRANSMEM 747 768
CC DOMAIN 769 781
CC TRANSMEM 782 803
CC DOMAIN 804 818
CC TRANSMEM 819 843
CC DOMAIN 844 908
CC CARBOHYD 909 955
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC
CC -----
seq_name: SwissProt_39:UBP2_KLUJA
seq_documentation_block:
ID UB2_KLUJA STANDARD; PRT; 1221 AA.
AC AC
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 2 (EC 3.1.2.15) (UBIQUITIN
DE THIOLESTERASE 2) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 2)
DE (DEUBIQUITINATING ENZYME 2).
DE UB2.
CC Kluyveromyces lactis (Yeast).
CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
CC NCBI_TaxID=28985;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=JBD100;
CC Winkler A.A.; Korstanje R.; Zonneveld B.J.M.; Steensma H.Y.;
CC "KlUp2, a ubiquitin hydrolase from Kluyveromyces lactis is one of
CC several ubiquitin-related genes that can complement a mutation in the
CC centromeric protein Cbf2 of Saccharomyces cerevisiae."
CC Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HAS AN ATP-INDEPENDENT ISOPEPTIDASE ACTIVITY, CLEAVING
CC AT THE CARBOXYL TERMINUS OF THE UBIQUITIN MOIETY IN NATURAL OR
CC ENGINEERED LINEAR FUSION PROTEINS, IRRESPECTIVE OF THEIR SIZE OR
CC THE PRESENCE OF AN AMINO-TERMINAL EXTENSION TO UBIQUITIN (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O =
CC UBIQUITIN + A THIOL.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19; ALSO KNOWN AS
CC FAMILY 2 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
EMBL; AF022776; AAB94074.1; .
CC
```

```
alignment_scores:
Quality: 65.00 Length: 44
Ratio: 2.167 Gaps: 1
Percent Similarity: 68.182 Percent Identity: 29.545
alignment_block:
US-09-049-696-14/rev x MGR8_RAT ..
Align_seg 1/1 to: MGR8_RAT from: 1 to: 908
131 GTCCTGAGATCAAGATACGTGTACT.....TATTCGAATGATATACT 88
744 IIESERASPLEUSERLEULCYSERLEUGLYTYRSERILEULEUW 760
87 TGTGAGCTGTTCCATGTCATATCATCCCGAGAGCTGTCCAGTCAGA 38
760 TValThrCysThrValTyraAlaIleLysThrArgGlyValProGluThrP 777
37 TTAATGACACTGCCCGCTGAATTTCCGCTT 6
777 heAsnGluAlaLysProIleGlyPheThrMet 787
seq_name: SwissProt_39:UBP2_KLUJA
seq_documentation_block:
ID UB2_KLUJA STANDARD; PRT; 1221 AA.
AC AC
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 2 (EC 3.1.2.15) (UBIQUITIN
DE THIOLESTERASE 2) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 2)
DE (DEUBIQUITINATING ENZYME 2).
DE UB2.
CC Kluyveromyces lactis (Yeast).
CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
CC NCBI_TaxID=28985;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=JBD100;
CC Winkler A.A.; Korstanje R.; Zonneveld B.J.M.; Steensma H.Y.;
CC "KlUp2, a ubiquitin hydrolase from Kluyveromyces lactis is one of
CC several ubiquitin-related genes that can complement a mutation in the
CC centromeric protein Cbf2 of Saccharomyces cerevisiae."
CC Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HAS AN ATP-INDEPENDENT ISOPEPTIDASE ACTIVITY, CLEAVING
CC AT THE CARBOXYL TERMINUS OF THE UBIQUITIN MOIETY IN NATURAL OR
CC ENGINEERED LINEAR FUSION PROTEINS, IRRESPECTIVE OF THEIR SIZE OR
CC THE PRESENCE OF AN AMINO-TERMINAL EXTENSION TO UBIQUITIN (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O =
CC UBIQUITIN + A THIOL.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19; ALSO KNOWN AS
CC FAMILY 2 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
EMBL; AF022776; AAB94074.1; .
CC
```


DR	1	EFAM: PF02472; Exbd; 1.
KW	1	Transport; Protein transport; Transmembrane; Inner membrane.
ET	9	CYTOPLASMIC (POTENTIAL).
ET	10	POTENTIAL.
ET	31	POTENTIAL.
ET	15	PERIPLASMIC (POTENTIAL).
SQ	145	AA; 15728 MW; D43D7E267C20CB45 CRC64; SEQUENCE

alignment_scores:		Length:
Quality:	64.50	97
Ratio:	1.217	Gaps: 5
Similarity:	54.639	Percent Identity: 23.711

alignment_block: US-09-049-696-14/rev x EXBD_PASHA

Align: Seq 1/1 to: EXBD_PASHA from: 1 to: 145

248 AAGAATCTGGGCCATTTTCAAAAAGTAATGTTTTTCGGTTTAAACAAAAAGA 191
55 LysLeuLeuThrIleThrGluAsnAsnGluPhePheAsnAspGlnAl 71

198 CTTCCTCAGAGTTGGCTTCCTT.....TGGGATGAGAGCA 164

[illegible]

163 GTAGTATTCACCTGAAG.....AGATTCATT 138

88 InLysValSerLeuLysValAspGlyAlaValAlaPheGluLysPheVal 104

137 GAACTT.....GTCTCTGAGATCAAGAATACTTGTACTTA 103

```

. |||||
105 GluLeuThrAspIleLeuSerAlaAsnGluIleLysAsnValAla..... 119
      : : : : :

```

102 TTCCAATGATATACTTGTGAGCTGTTCCATGGTCATAATCATCCCCAGGA 53

120IleIleThrLysLysGluThrAlaProAlaProSers 132

52 GC.....TGTCCAAGTCAGATTAAATGAGACTGCCCCC 21
|||
132 erThrProGlySerProAlaGlnValProAlaValAlaPro 145
|||:::||||:::|:::|||||

This Page Blank (uspto)

OM of: US-09-049-696-14 to: PIR_58:* out_format : pfs

Date: Mar 30, 2002 2:27 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+_n2p.model -DEV=xlp
-Q/cgn2.1/USPTO_spool/US09049696/runat_280332002_145238_2031/app_query.fasta_1.12579
-B/Pir_58 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-OGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09049696 -CGN1_1_367 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-049-696-14
Query length: 248
Database: PIR_68:*
Database sequences: 219241
Database length: 76174552
Search time (sec): 470.790000
```

score.list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
pir2:JG0168	+	34.00	636.85	913	! gob-5 protein - mouse
pir2:T02205	+	195.00	399.87	3.1e-15	! Lu-ECAM-1 protein - bovine
pir2:R86336	+	73.00	139.14	1.73	! hypothetical protein AAF88159.1
pir2:B75569	+	72.50	140.66	1.91	! conserved hypothetical protein
pir2:A59235	+	69.00	120.11	6.12	! unconventional myosin heavy ch
pir2:T31610	+	68.50	127.36	6.24	! hypothetical protein Y50E8A.f
pir2:E64545	+	68.00	124.83	7.32	! hypothetical protein HP0205 - H
pir2:T42634	+	68.00	122.61	7.56	! connectin/titin - chicken (fra
pir1:GNNTN	+	68.00	115.46	8.39	! genome polyprotein - murine po
pir2:T32320	+	67.50	131.70	7.50	! hypothetical protein M11348 - M
pir2:C64468	+	67.50	131.30	7.54	! hypothetical protein M11348 - M
pir2:T05511	+	67.00	129.27	7.77	! hypothetical protein F13M32.40
pir2:RA0692	+	67.00	124.13	9.47	! signal recognition particle 72K
pir2:T14597	+	66.50	129.50	9.91	! proteinase homolog - slime mold
pir2:E64395	+	66.50	124.82	10.61	! hybrid cluster [4Fe-2S-3O] pro
pir2:T18289	+	66.50	116.99	11.89	! racGAP protein - slime mold (D
pir2:G71088	+	66.00	129.06	11.29	! hypothetical protein PH0972 - H
pir2:T37398	+	66.00	117.34	13.39	! hypothetical protein M18.5 - C
pir2:T22574	+	65.50	125.26	13.50	! hypothetical protein F53P4.9 -
pir2:T16284	+	65.00	130.42	14.17	! hypothetical protein F35D11.7 -
pir2:T49142	+	65.00	117.12	17.20	! metabotropic glutamate recepto
pir2:T30529	+	65.00	114.51	17.86	! ubiquitin carboxyl-terminal hy
pir2:S78055	+	64.50	132.16	15.64	! Ig heavy chain precursor V-D-J
pir2:T39341	+	64.00	127.61	18.91	! hypothetical protein SPB1271.0
pir2:S67380	+	64.00	125.16	19.59	! probable cell-wall synthesis pr
pir2:D65210	+	64.00	117.26	21.98	! yjBH protein precursor - Escher
pir2:JH0563	+	64.00	114.90	22.75	! metabotropic glutamate recepto
pir2:A49874	+	64.00	114.87	22.76	! metabotropic glutamate recepto
pir2:T38346	+	64.00	95.85	30.02	! elastic titin - human (fragmen
pir2:D82089	+	63.50	133.08	19.76	! conserved hypothetical protein
pir2:JE0316	+	63.50	127.25	21.51	! protozoan RNase/Ddi - slime mol
pir2:T39993	+	63.50	119.48	24.08	! 6-phospho-beta-glucosidase (EC
pir2:T71048	+	63.50	119.44	24.09	! hypothetical protein PH1678 - F
pir2:T25162	+	63.50	117.98	24.61	! Frizzled-1 protein homolog - Ca
pir2:B61190	+	63.00	124.29	25.41	! mast cell growth factor, short
pir2:T11972	+	63.00	124.15	25.46	! hypothetical protein ORF249 - H
pir2:A35974	+	63.00	123.34	25.76	! mast cell growth factor precurs
pir2:EB4806	+	63.00	120.48	26.86	! probable elongation factor (imp
pir2:T32449	+	63.00	112.45	30.18	! hypothetical protein F52D1.1 -
pir1:GNVTP	+	63.00	104.58	33.85	! genome polyprotein - murine po
pir2:SL3554	+	63.00	104.58	33.85	! genome polyprotein - murine po

```
pir2:T09636 - 62.50 122.75 29.41 258 ! ycsE protein - Lactobacillus
pir2:T22908 - 62.50 116.17 32.36 545 ! hypothetical protein F58D2.1
pir2:T33823 + 62.00 119.58 34.85 327 ! hypothetical protein W05F2.5
pir2:B81832 + 62.00 118.25 35.52 380 ! probable polyamine permease

seq_name: pir2:JG0168

seq_documentation_block:
gob-5 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
C:Accession: JG0168
R:Komiyama, T.; Tanigawa, Y.; Hirohashi, S.
Biochem. Biophys. Res. Commun. 255, 347-351, 1999
A:Title: Cloning and identification of the gene gob-5, which is expressed in intestine
A:Accession: JG0168
A:Reference number: JG0168; MUID:99160866
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-913 <KOM>
A:Cross-references: DDBJ:AB016592
```

```
alignment_scores:
Quality: 304.00 Length: 82
Ratio: 4.222 Gaps: 0
Percent Similarity: 87.805 Percent Identity: 71.951
```

alignment_block:

```
US-09-049-696-14 x JG0168
```

```
Align seg 1/1 to: JG0168 from: 1 to: 913
```

```
3 CTGAAGCGGAAATTCACGGGGGAGTCATTCATTCGACTTGACAGC 52
759 LeulysAlaSerIleGlnGlnAsnLeuValAsnLeuThrTrpAl 775
53 TCCTGGGATGATTCATGACCTGACAGCTCAGCTCAGCTATCATTCGAA 102
775 aProGlyAspAspTyrAspHisGlyArgAlaSerAsnTyrIleIleArgM 792
103 TAAGTACAGTATTCCTTCATTCAGACACAAAGTTCATCAATGAATCTTCAA 152
792 etSerThrSerIleValAspLeuArgAspHisPheAsnThrSerLeuGln 808
153 GTGAATACCTACTGCTCTCATCCCAAGGAAGCAACTCTGAGGAGTCTT 202
809 ValAsnThrThrGlyLeuIleProLysGluAlaSerSerGluGluIlePh 825
203 TTGCTTTAAACCAAGAAACATTTCTTTGAAAATGGCAGACATCTT 248
825 eGluPheGluLeuGlyGlyAsnThrPheGlyAsnGlyThrAspIle 840
```

```
seq_name: pir2:T02205
```

seq_documentation_block:

```
Lu-ECAM-1 protein - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 01-Dec-2000
C:Accession: T02205; T02152; T02171
R:Elble, R.C.; Widom, J.; Gruber, A.D.; Abdel-Chay, M.; Levine, R.; Goodwin, A.; Pau
submitted to the EMBL Data Library, April 1997
A:Description: Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial
A:Reference number: Z14590
A:Accession: T02205
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-905 <ELB>
A:Cross-references: EMBL:AF001261; NID:Q2623762; PIDN:AAB86529.1; PID:G2623763
A:Experimental source: lung
A:Accession: T02152
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
```


[illegible]

C;Accession: T32320

C;Accession: T32320

1000

70 ValTrpCysValPheCysAspGluAspIle 79

seq_name: pir2:A40692

seq_documentation_block:

signal recognition particle 72k chain - dog
C:Species: Canis lupus familiaris (dog)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999C:Accession: A40692; S32167
R:Lutcke, H.; Prehn, S.; Ashford, A.J.; Remus, M.; Frank, R.; Dobberstein, B.
J. Cell Biol. 121, 977-985, 1993
A:Title: Assembly of the 68- and 72-kD proteins of signal recognition particle with 7S R
A:Reference number: A40692; MUID:93273803

A:Accession: A40692

A:Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-671 <LUT>

A:Cross-references: EMBL:X67813; NID:g297767; PIDN:CA448014.1; PID:g297768

A:Note: sequence extracted from NCBI backbone (NCBIP:132901)

alignment_scores:	Quality:	67.00	Length:	78
	Ratio:	1.457	Gaps:	4
Percent Similarity:	58.974	Percent Identity:	26.923	

alignment_block:

US-09-049-696-14 x A40692 ..

Align seg 1/1 to: A40692 from: 1 to: 671

12 GAAATTCACGGGGGAGCTCTCAATATCGACTTGGACAGCTCTGGGGA 61

|||||:||||| ||| :|||||

113 GluLeuTyrglyGlnValLeuTyArgLeu.....G1 123

62 TGATTATGACCATGGAACGCT...CACAGATATCATTCGAATAAGTA 108

: |||||:||||| ||| :|||||

123 uArgTyArgPglucysLeuAlaValTyArgAspLeuValArgAsnSer. 139

109 CAAGTATCTTGATCTCAGACAGCAAGTTCAATGAATCTCTCAAGTGAAT 158

: |||||:||||| ||| :|||||

140GlnAspTyArgAspGluArgLysThrAsn 150

159 ACTACTGCTCTCATCCCAAGGAAGCAACTCTGAGGAAGTCTTTTGT 208

: |||||:||||| ||| :|||||

151 LeuSerAlaValAlaAlaAlaGlnSerAsnTrpGluLysValVal..... 165

209 TAAACCAAGAAACATTAATCTTTTGAATGGCACA 242

|||||:||||| ||| :|||||

166ProGluAsnLeuGlyLeuGlnGluGlyThr 175

seq_name: pir2:T14597

seq_documentation_block:

proteinase homolog - slime mold (Dictyostelium discoideum) retrotransposon Skipper (frag
C:Species: Dictyostelium discoideum

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000

C:Accession: T14597

R:Leng, P.; Klatte, D.H.; Schumann, G.; Boeke, J.D.; Steck, T.L.

Nucleic Acids Res. 26, 2008-2015, 1998

A:Title: Skipper, an LTR retrotransposon of Dictyostelium.

A:Reference number: Z18156; MUID:98186844

A:Accession: T14597

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-322 <LEN>

A:Cross-references: EMBL:AF049230; NID:g3135476; PID:g3135478; PIDN:AAC39020.1

C:Genetics:

A:Gene: pro

A:Mobile element: retrotransposon Skipper

alignment_scores:

Quality:	66.50	Length:	55
Ratio:	1.750	Gaps:	2
Percent Similarity:	69.091	Percent Identity:	32.727

alignment_block:

US-09-049-696-14 x T14597 ..

Align seg 1/1 to: T14597 from: 1 to: 322

84 CACAAGTATATCATCGAATAAGTACAAGTATCTTGATCTCAGACAAA 133

||||| ||| :|||||

61 HisLysAsnIleAsnLysLeuLysGluTyArgVal.....LysaspSe 74

134 GTTCAATCAATCTCTTCAAGTGAATACTACTGCTCTCTCATCCCAAGGA 183

: |||||:||||| ||| :|||||

74 rTyrglyIleLysLeuValProAspThrValProIleProLysAlav 91

184 CCAACTCTGAGGAAGTCTTTTGTAAACGACAAACATTAATCTTTGAA 233

: |||||:||||| ||| :|||||

91 alGluSerGlu.....MetLysTyrlsProGlnLysIleGluPheAsn 105

234 AATGGCACAGATCTT 248

: |||||:|||||

106 ThrProThrGluLeu 110

seq_name: pir2:E64395

seq_documentation_block:

hybrid cluster [4Fe-2S-3O] protein MJ0765 [similarity] - Methanococcus jannaschii

N:Alternate names: prismane [6Fe-6S] protein [mismomer]

C:Species: Methanococcus jannaschii

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: E64395

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak

; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,

rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc

A:Reference number: A64300; MUID:96337999

A:Accession: E64395

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-548 <BUL>

A:Cross-references: GB:067522; GB:L77117; NID:g2826315; PIDN:AAB98761.1; PID:g1592305

C:Genetics:

A:Map position: REV688457-686811

C:Superfamily: Desulfovibrio hybrid cluster [4Fe-2S-3O] protein; hybrid cluster [4Fe-

C:Keywords: 4Fe-2S-3O; 4Fe-4S; electron transfer; iron-sulfur protein; metalloprotein

F:221-513/Domain: hybrid cluster [4Fe-2S-3O] homology <HCL>

F:12,15,24,30/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

F:244,268,312,403,431,456,491/Binding site: 4Fe-2S-3O cluster (His, Glu, Cys, Cys, Cy

F:403/Modified site: cysteine persulfide (Cys) #status predicted

alignment_scores:

Quality: 66.50

Ratio: 2.078

Percent Similarity: 62.745

Percent Identity: 37.255

alignment_block:

US-09-049-696-14 x E64395 ..

Align seg 1/1 to: E64395 from: 1 to: 548

66 TATGACCATGGACAGCTCACAGTATATATCATTCGAATAAGTACAAGTAT 115

|||||:||||| |||||:|||||

170 TyrAsnAsnGluAspIleHisLysPheIleLysAlaPheThrLysI 186

115 TCTTGATCTCAGACACAAG.....TTCAATGAATCTCTTCAA.... 152

: |||||:||||| |||||:|||||

186 eValAspSerLysAspAlaAspGluLeuPheAsnLeuAlaMetGluThrG 203


```
153 .....GTGAATACTACTGCTCTCATCCCAAGGAGCCCAACTCT 191
||||:||||| |||||:| | |||||:| |
203 lYlYsTyrAlaValGluThrLeuAlaLeuLeuAspLys...AlaAsnThr 218
192 GAG 194
|||
219 Glu 219
```

This Page Blank (uspto)

153 ValAsnThrThrAlaLeuIleProLysGluAlaAsnSerGluGluValPh 169

203 TTGTGTTTAAACGAGAAACATTACTTTTGAATAATGCACAGATCTT 248
|||||
169 eLeuPheLysProGluAsnIleThrPheGluAsnGlyThrAspLeu 184

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US95-07289-9

seq_documentation_block:
; Sequence 9, Application PC/TUS9507289
; GENERAL INFORMATION:
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Rosen, Craig
; TITLE OF INVENTION: Colon Specific Genes and Proteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; ADDRESSEE: Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07289
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Perraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-265
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-07289-9

alignment_scores:
Quality: 427.00 Length: 82
Ratio: 5.207 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-14 x PCT-US95-07289-9 ..
Align seg 1/1 to: PCT-US95-07289-9 from: 1 to: 228
3 CTGAAGCGGAAATTCACGGGGCAGTCTCATTAACTGACTTGGACAGC 52
|||||
103 LeuLysAlaGluIleHisGlySerLeuIleAsnLeuThrTrpThrAl 119
53 TCCTGGGGATGATTATGACCATGGAACAGCTCACAAAGTATATATCGAA 102
|||||
119 aProGlyAspAspTyrAspHisGlyThrAlaHisLysTyrIleIlearg 136
103 TAAGTACAAGTATTCTTGATCTCAGACACAAGTTCAATGAATCTCTTCAA 152
|||||
136 leSerThrSerIleLeuAspLeuargAspLysPheAsnGluSerLeuGln 152
153 GTGAATACTACTGCTCTCATCCCAAGGAAGCCCAACTCTCTGAGGAAGTCTT 202
|||||
153 ValAsnThrThrAlaLeuIleProLysGluAlaAsnSerGluGluValPh 169

203 TTGTGTTTAAACGAGAAACATTACTTTTGAATAATGCACAGATCTT 248
|||||
169 eLeuPheLysProGluAsnIleThrPheGluAsnGlyThrAspLeu 184

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-726-227-1

seq_documentation_block:
; Sequence 1, Application US/08726227
; Patent No. 5780235
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL HUMAN VOLTAGE-DEPENDENT
; TITLE OF INVENTION: ANION CHANNEL HACH
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,227
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0135 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
US-08-726-227-1

alignment_scores:
Quality: 69.00 Length: 62
Ratio: 2.156 Gaps: 2
Percent Similarity: 51.613 Percent Identity: 32.258
alignment_block:
US-09-049-696-14 x US-08-726-227-1 ..
Align seg 1/1 to: US-08-726-227-1 from: 1 to: 283

21 GGGGCGAGTCTC.....ATTAA 37
|||||
191 GlyGlySerIleTyrGlnLysValAsnGluLysIleGluThrSerIleAs 207
38 TCTGACTTCGACAGCTCCTGGGATGATGACCATGGACAGCTCACA 87
|||||
207 nLeuAlaIatrpThrAlaGlySerAsnAsnThrArgPheGlyIleAlaLal 224
88 AGTATATCATTCGAATAAGTACAAAGTATTCTTGATCTCTCAGACAAAGTTC 137

|||||:..... |||||:..... 233
224 ystYrmetLeuAspCysArgThrSerLeu..... 233
138 AATGAATCTCTTCAAGTAAGTACTACTCTCATC 173
||| :|||||:.....:|||||
234SerAlaLysValAsnAsnAlaSerLeulle 243

seq_name: /cgn2_5/ptodata/2/iaa/6A_COMB.pep:US-08-826-246-6

seq_documentation_block:
; Sequence 6, Application US/08826246
; Patent No. 6048709
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/826,246
; FILING DATE: 28-MAR-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE: 13-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/011,787
; FILING DATE: 16-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-078-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-826-246-6

alignment_scores:
Quality: 69.00 Length: 62
Ratio: 2.156 Gaps: 2
Percent Similarity: 51.613 Percent Identity: 32.258

alignment_block:

US-09-049-696-14 x US-08-826-246-6

Align seg 1/1 to: US-08-826-246-6 from: 1 to: 283

21 GGGGCGAGCTCTC.....ATTAA 37
|||||:.....
191 GlyGlySerIleTyrGlnLysValAsnGluLysThrSerIleAs 207

38 TCTGACITGGACAGCTCTCTGGGATGATTATGACCATGGACAGCTCACA 87
|||||:|||||:.....:|||||:.....
207 nLeuAlaIrpThrAlaGlySerAsnAsnThrArgPheGlyIleAlaAlaL 224
88 AGTATATCATTCGAATAAGTACAAGTATTCTTGATCTCAGACAGAGTTC 137
|||||:.....:|||||:.....
224 ystYrmetLeuAspCysArgThrSerLeu..... 233
138 AATGAATCTCTTCAAGTAAGTACTACTCTCTCATC 173
||| :|||||:.....:|||||
234SerAlaLysValAsnAsnAlaSerLeulle 243

seq_name: /cgn2_5/ptodata/2/iaa/6A_COMB.pep:US-08-944-495-6

seq_documentation_block:
; Sequence 6, Application US/08944495
; Patent No. 6087477
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,495
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-067-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 283 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-944-495-6

alignment_scores:
Quality: 69.00 Length: 62
Ratio: 2.156 Gaps: 2
Percent Similarity: 51.613 Percent Identity: 32.258

alignment_block:

US-09-049-696-14 x US-08-944-495-6

Align seg 1/1 to: US-08-944-495-6 from: 1 to: 283

21 GGGGCGAGCTCTC.....ATTAA 37
|||||:.....


```

178 AGGAAGCCAACTCTAGGAAGCTTTTGTGTTAAACAGAA 218
A |||:: :||| ||| ||||| |||||
101 ysSerPheLySerProGluProArgLeuPheThrProGlu 114

seq_name: /cnp2_6/ptodata/2/iaa/5A_COMB.pep.US-08-726-227-4
seq_documentation_block:
  Sequence 4, Application US/08726227
  Patent No. 5780235
  GENERAL INFORMATION:
    APPLICANT: Bandman, Olga
    APPLICANT: Hillman, Jennifer L.
    TITLE OF INVENTION: NOVEL HUMAN VOLTAGE-DEPENDENT
    TITLE OF INVENTION: ANION CHANNEL HACH
    NUMBER OF SEQUENCES: 4
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Incyte Pharmaceuticals, Inc.
      STREET: 3174 Porter Drive
      CITY: Palo Alto
      STATE: CA
      COUNTRY: U.S.
      ZIP: 94304
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Diskette
      COMPUTER: IBM Compatible
      OPERATING SYSTEM: DOS
      SOFTWARE: FastSeq Version 1.5
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/726,227
      FILING DATE: Filed Herewith
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:
      FILING DATE:
    ATTORNEY/AGENT INFORMATION:
      NAME: Billings, Lucy J.
      REGISTRATION NUMBER: 36,749
      REFERENCE/DOCKET NUMBER: PF-0135 US
      TELECOMMUNICATION INFORMATION:
        TELEPHONE: 415-855-0555
        TELEFAX: 415-845-4166
      INFORMATION FOR SEQ ID NO: 4.

```

```

? LENGTH: 283 amino acids
?
? .TYPE: amino acid
?
? STRANDEDNESS: single
?
? TOPOLOGY: linear
?
? MOLECULE TYPE: peptide
?
```

```
:  
HYPOTHETICAL: NO  
:  
ANTI-SENSE: NO  
:  
FRAGMENT TYPE: N-terminal
```

```

; ORIGINAL SOURCE:
; IMMEDIATE SOURCE:

```

```

; LIBRARY: GenBank
; CLONE: 1463028
US-08-726-227-4

```

alignment_scores:		
Quality:	65.00	Length: 62
Ratio:	2.097	Gaps: 2

alignment_block:
US-05-049-696-14 x US-08-736-227-4
Percent Similarity: 50.000 Percent Identity: 32.250

Align seg 1/1 to: US-08-726-227-4 from: 1 to: 283

21 GGGGGCAGTCTC.....ATTAA 37

191 GlvGlySerIleTyrGlnLysValAsnGluAraGluGluThrSerIleAs 207

38 TCTGACTTGGACAGCTCCTGGGGATGATTATGACCATGGAACAGCTCACA 87

[illegible]

207 nLeuAlaTrpThrAlaGlySerAsnAsnThrArgPheGlyIleAlaAlaL 224
88 AGTATATCATTCGATTAAGTACAAAGTATCTTGATCTCAGACACAGTTC 137
224 ystYrLysLeuAspCysArgThrSerLeu..... 233
138 AATGAATCTCTTCAAGTGAATCACTACTGCTCTCATC 173
234SerAlaLysValAsnAsnAlaSerLeuLe 243
seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-823-110-1

seq_documentation_block:
; Sequence 1, Application US/08823110
; Patent No. 6077675
; GENERAL INFORMATION:
; APPLICANT: Stormann, Thomas M.
; APPLICANT: Simin, Rachel T.
; APPLICANT: Hammerland, Lance G.
; APPLICANT: Fuller, Forrest H.
; TITLE OF INVENTION: NOVEL HUMAN METABOTROPIC
; TITLE OF INVENTION: GLUTAMATE RECEPTOR
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/823,110
; FILING DATE: March 24, 1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/604,298
; FILING DATE: February 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 224/259
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 908 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Peptide
US-08-823-110-1

alignment_scores:
Quality: 65.00 Length: 44
Ratio: 2.167 Gaps: 1
Percent Similarity: 68.182 Percent Identity: 29.545

alignment_block:

US-09-049-696-14/rev x US-08-823-110-1 ..

Align seg 1/1 to: US-08-823-110-1 from: 1 to: 908

131 GTCCTGAGATCAAGAACTGTACT.....TATTCGAATGATATACT 88

744 IleSerAspLeuSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 760
87 TGTGAGCTGTTCCATGGTCATAATCCCGAGGAGCTGTCCAAAGTCAGA 38
760 ValThrCysThrValYrAlaIleLysThrArgGlyValProGluThrP 777
137 TTAATGAGACTGCCCGCTGAATTCGCCGCTT 6
777 heAsnGluAlaLysProIleGlyPheThrMet 787
seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-604-298-1

seq_documentation_block:
; Sequence 1, Application US/08604298
; Patent No. 6084084
; GENERAL INFORMATION:
; APPLICANT: Stormann, Thomas M.
; APPLICANT: Simin, Rachel T.
; APPLICANT: Hammerland, Lance G.
; APPLICANT: Fuller, Forrest H.
; TITLE OF INVENTION: NOVEL HUMAN METABOTROPIC
; TITLE OF INVENTION: GLUTAMATE RECEPTOR
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/604,298
; FILING DATE: February 21, 1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 212/044
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 908 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Peptide
US-08-604-298-1

alignment_scores:
Quality: 65.00 Length: 44
Ratio: 2.167 Gaps: 1
Percent Similarity: 68.182 Percent Identity: 29.545

alignment_block:

US-09-049-696-14/rev x US-08-604-298-1 ..

Align seg 1/1 to: US-08-604-298-1 from: 1 to: 908

3 CTGAAGCGGAAATTCACGGGGCAGTCTCATTTAATCTGACTTGACAGC 52
||| ||| :::: :::: ||| :: ::
19 LeuValAlaAsnLeuProLysAspCysMetIleThrLeuLysTyr...Va 34

53 TCCTCGGGATGATTAT.....GACCATGGAAACAGCTCACAAAGTATATCA 96

359 heAsnGluAlaLysProIleGlyPheThrMet 369

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-617-785-4

seq_documentation_block:
; Sequence 4, Application US/08617785E
; Patent No. 6228610
; GENERAL INFORMATION:
; APPLICANT: Flor, Peter J.
; APPLICANT: Kuhn, Ranier
; APPLICANT: Lindaur, Kristen
; APPLICANT: Puttner, Irene
; APPLICANT: Knopfel, Thomas
; TITLE OF INVENTION: Human Metabotropic Glutamate Receptor Subtypes (HMR4,
; TITLE OF INVENTION: HMR6, HMR7) and Related DNA Compounds
; FILE REFERENCE: 4-19679/A/PCT
; CURRENT APPLICATION NUMBER: US/08/617,785E
; CURRENT FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: PCT/EP94/02991
; EARLIER FILING DATE: 1994-09-07
; EARLIER APPLICATION NUMBER: EPO 9416553.7
; EARLIER FILING DATE: 1994-08-19
; EARLIER APPLICATION NUMBER: EPO 93810663.0
; EARLIER FILING DATE: 1993-09-20
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 867
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-617-785-4

alignment_scores:
Quality: 64.00 Length: 44
Ratio: 2.065 Gaps: 1
Percent Similarity: 70.455 Percent Identity: 27.273

alignment_block:
US-09-049-696-14/rev x US-08-617-785-4 ..
Align seg 1/1 to: US-08-617-785-4 from: 1 to: 867

131 GTCTCTGAGATCAAGATCTGTACT.....TATTCGAATGATATACT 88
696 ILeThrAspLeuGlnIleIleCysSerLeuGlyTyrSerIleLeuLeuMe 712
87 TGTGAGCTGTTCCATGTCATATCATCCCGAGGAGCTGCCAAGTCAGA 38
712 tValThrCysThrValTyrAlaIleLysThrArgGlyValProGluAsn 729
37 TTAATGAGACTGCCCGCGTGAATTTCCGCCTT 6
729 heAsnGluAlaLysProIleGlyPheThrMet 739

seq_documentation_block:
; Sequence 2, Application US/08617785E
; Patent No. 6228610
; GENERAL INFORMATION:
; APPLICANT: Flor, Peter J.
; APPLICANT: Kuhn, Ranier
; APPLICANT: Lindaur, Kristen
; APPLICANT: Puttner, Irene
; APPLICANT: Knopfel, Thomas
; TITLE OF INVENTION: Human Metabotropic Glutamate Receptor Subtypes (HMR4,
; TITLE OF INVENTION: HMR6, HMR7) and Related DNA Compounds
; FILE REFERENCE: 4-19679/A/PCT
; CURRENT APPLICATION NUMBER: US/08/617,785E
; CURRENT FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: PCT/EP94/02991
; EARLIER FILING DATE: 1994-09-07
; EARLIER APPLICATION NUMBER: EPO 9416553.7
; EARLIER FILING DATE: 1994-08-19
; EARLIER APPLICATION NUMBER: EPO 93810663.0
; EARLIER FILING DATE: 1993-09-20
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 867
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-617-785-4

alignment_scores:
Quality: 64.00 Length: 44
Ratio: 2.065 Gaps: 1
Percent Similarity: 70.455 Percent Identity: 27.273

alignment_block:
US-09-049-696-14/rev x US-08-617-785-4 ..
Align seg 1/1 to: US-08-617-785-4 from: 1 to: 867

131 GTCTCTGAGATCAAGATCTGTACT.....TATTCGAATGATATACT 88
696 ILeThrAspLeuGlnIleIleCysSerLeuGlyTyrSerIleLeuLeuMe 712
87 TGTGAGCTGTTCCATGTCATATCATCCCGAGGAGCTGCCAAGTCAGA 38
712 tValThrCysThrValTyrAlaIleLysThrArgGlyValProGluAsn 729
37 TTAATGAGACTGCCCGCGTGAATTTCCGCCTT 6
729 heAsnGluAlaLysProIleGlyPheThrMet 739

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-617-785-2

seq_documentation_block:
; Sequence 2, Application US/08617785E
; Patent No. 6228610
; GENERAL INFORMATION:
; APPLICANT: Flor, Peter J.
; APPLICANT: Kuhn, Ranier
; APPLICANT: Lindaur, Kristen
; APPLICANT: Puttner, Irene
; APPLICANT: Knopfel, Thomas
; TITLE OF INVENTION: Human Metabotropic Glutamate Receptor Subtypes (HMR4,
; TITLE OF INVENTION: HMR6, HMR7) and Related DNA Compounds
; FILE REFERENCE: 4-19679/A/PCT
; CURRENT APPLICATION NUMBER: US/08/617,785E
; CURRENT FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: PCT/EP94/02991
; EARLIER FILING DATE: 1994-09-07
; EARLIER APPLICATION NUMBER: EPO 9416553.7
; EARLIER FILING DATE: 1994-08-19
; EARLIER APPLICATION NUMBER: EPO 93810663.0

OM of: US-09-049-696-14 to: A_Geneseq_1101.* out_format : pfs
Date: Mar 30, 2002 2:18 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+g2p_model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09049696/runat_28032002_145237_2001/app_query.fasta_1.12579
-DB=A_Geneseq_1101 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-XGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1.293 -NCFU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-049-696-14
Query length: 248
Database: A_Geneseq_1101.*
Database sequences: 522463
Database length: 74073290
Search time (sec): 819.910000

score_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
/SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT:AAW06548 +				228	4.2e-49
/SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW46879 +				228	4.27.00 1018.35
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW75614 +				869	4.27.00 1018.35
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW74824 +				914	4.27.00 1004.34
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW24514 +				914	4.27.00 1003.81
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW73716 +				914	4.27.00 1003.81
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW74733 +				914	4.27.00 1003.81
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW75474 +				925	4.27.00 1003.68
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW73854 +				552	4.21.00 994.51
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW74822 +				913	3.0e-32
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW73715 +				913	3.04.00 704.96
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW66749 +				919	259.50 596.77
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW87560 +				919	259.50 596.77
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW65272 +				919	259.50 596.77
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW45904 +				942	159.00 354.82
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW70459 +				942	159.00 352.32
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW74823 +				943	159.00 352.31
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW41036 +				943	159.00 352.31
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW11321 +				943	159.00 352.31
/SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT:AAW36004 +				283	69.00 146.24
/SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW48908 +				283	69.00 146.24
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW07222 +				283	69.00 146.24
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW45015 +				283	69.00 146.24
/SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW40538 +				1099	66.50 125.96
/SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:AAW14770 +				166	66.00 144.54
/SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAW4608 +				469	66.00 137.66
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW49114 +				260	65.00 137.41
/SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW4928 +				908	65.00 124.31
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW49130 +				1059	65.00 122.70
/SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAW49131 +				1418	65.00 119.65
/SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:AAW14766 +				166	64.00 139.68
/SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:AAW72095 +				481	64.00 128.54
/SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:AAW72093 +				867	64.00 122.37
/SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:AAW82658 +				912	64.00 121.84
/SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:AAW80479 +				915	64.00 121.80
/SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:AAW72097 +				915	64.00 121.80
/SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:AAW72098 +				922	64.00 121.72
/SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:AAW73270 +				141	63.00 138.96
/SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT:AAW32167 +				155	63.00 137.97
/SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT:AAW32165 +				165	63.00 137.31

/SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:AAW14761 + 63.00 137.31 6.83
/SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW77781 + 63.00 137.31 6.83
/SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW51359 + 63.00 137.31 6.83
/SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:AAW14771 + 63.00 137.25 6.85
/SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:AAW83789 + 63.00 137.25 6.85

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT:AAW06548

seq_documentation_block:

ID AAW06548 standard; Protein; 228 AA.
XX
AC AAW06548;
XX
DT 13-MAR-1997 (first entry)
DE Human colon specific gene CSG5 polypeptide fragment.
DE
KW Colon specific gene; CSG5; colon cancer; metastasis; diagnosis;
KW therapy; antibody; vaccine.
XX
OS Homo sapiens.
XX
PN WO9639419-A1.
XX
PD 12-DEC-1996.
XX
PF 06-JUN-1995; 95WO-US07289.
XX
PR 06-JUN-1995; 95WO-US07289.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX

FI Rosen CA, Yu G;
XX
XX WPI; 1997-043054/04.
DR N-PSDB; AAT45884.
XX
XX
PT Human colon specific genes and their expression products - detection
PT of which, in non-colon tissue samples, can be used as indication of
PT colon cancer metastasis
XX
PS Claim 8; Fig 5; 60pp; English.
XX
XX Novel polypeptides (AAW06545-53) are encoded by cDNA clones (see also
CC AAT45880-92) corresponding to 13 human colon specific genes,
CC designated CSG1, CSG2, etc., that are primarily expressed in
CC tissues derived from the colon. Recombinant CSG polypeptides can
CC be produced in transformed host cells. They are useful diagnostic
CC markers for colon cancer and for colon cancer metastasis and can
CC also be used to screen for (ant)agonist cpds. of therapeutic or
CC diagnostic value. Antibodies raised against the colon-specific
CC polypeptides may be used to target colon cancer cells or as part
CC of a colon cancer vaccine.
XX
SQ Sequence 228 AA;

alignment_scores:

Quality: 427.00 Length: 82
Ratio: 5.207 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-14 x AAW06548 ..

Align seg 1/1 to: AAW06548 from: 1 to: 228

3 CTGAAGCGGGAATTCACGGGGCAGTCTCAATTAATCTGACTTGACAGC 52

103 LeuLysAlaGluIleHisGlySerLeuIleAsnLeuThrTrpThrAl 119

53 TCCTGGGAGTATTGACCATGGAACAGCTCAACAGTATATCATTCGAA 102

|||||

119 aProGlyAspAspTyrAspHisGlyThrAlaHisLysTyrIleIleArgI 136
 103 TAAGTACAGTATTCTTGATCTCAGAGACAAGTTCATGAATCTCTTCAA 152
 136 leSerThrSerIleLeuAspLeuArgAspLysPheAsnGluSerLeuGln 152
 153 GTGAATACTACTGCTCTCATCCCAAGGAAGCAACTCTCAGAGAACTCTT 202
 153 ValAsnThrAlaLeuIleProLysGluAlaAsnSerGluGluValPh 169
 203 TTGTTTAAACCAAGAAACATTACTTTTGAATAATGGCACAGATCTT 248
 169 eLeuPheLysProGluAsnIleThrPheGluAsnGlyThraspLeu 184

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:AAW46879

seq_documentation_block:
 ID AAW46879 standard; Protein; 228 AA.

XX AC AAW46879;

XX DT 22-JUN-1998 (first entry)

XX DE Protein sequence encoded by a colon-specific gene.

XX KW Colon-specific gene; probe; detection; expression; human;

XX KW diagnostic assay; colon cancer; antibody; screening.

XX OS Homo sapiens.

XX XX US573748-A.

XX PD 31-MAR-1998.

XX PF 06-JUN-1995; 95US-0469667.

XX PR 06-JUN-1995; 95US-0469667.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX XX Rosen C, Yu G;

XX DR WPI; 1998-229823/20.

XX DR N-PSDB; AAV16672.

XX PT Colon-specific nucleic acids - useful as probes for detecting colon cancer micrometastases

XX PS Claim 1; Fig 5; 5lpp; English.

XX CC AAW46876-80 and W4682-85 represent proteins encoded by colon-specific genes. The polynucleotides encoding these proteins can be used as probes to detect expression of the corresponding human genes, e.g. in diagnostic assays for detecting micrometastases of colon cancer. Recombinant cells containing the polynucleotides can be used to produce the proteins, in order that antibodies can be raised and used in further screening or diagnostics.

XX SQ Sequence 228 AA;

alignment_scores:
 Quality: 427.00 Length: 82
 Ratio: 5.207 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-049-696-14 x AAW46879 ..

Align seg 1/1 to: AAW46879 from: 1 to: 228

3 CTGAAGCGGAATTCACGGGGGCGAGTCTCAATTAATCTGACATGGACAGC 52
 ||||||||||||||||||||||||||||||||||||||||||||||||||||

103 LeuLysAlaGluIleHisGlySerLeuIleAsnLeuThrTrpThrAl 119
 53 TCCTGGGATGATTATGACCATGGAACAGCTCACAGTATATCATTCGAA 102
 119 aProGlyAspAspTyrAspHisGlyThrAlaHisLysTyrIleIleArgI 136
 103 TAAGTACAGTATTCTTGATCTCAGAGACAAGTTCATGAATCTCTTCAA 152
 136 leSerThrSerIleLeuAspLeuArgAspLysPheAsnGluSerLeuGln 152
 153 GTGAATACTACTGCTCTCATCCCAAGGAAGCAACTCTGAGGAAGTCTT 202
 153 ValAsnThrAlaLeuIleProLysGluAlaAsnSerGluGluValPh 169
 203 TTGTTTAAACCAAGAAACATTACTTTTGAATAATGGCACAGATCTT 248
 169 eLeuPheLysProGluAsnIleThrPheGluAsnGlyThraspLeu 184

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAG75614

seq_documentation_block:

ID AAG75614 standard; Protein; 869 AA.

XX AC AAG75614;

XX DT 03-SEP-2001 (first entry)

XX DE Human colon cancer antigen protein SEQ ID NO:6378.

XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection; colorectal carcinoma.

XX OS Homo sapiens.

XX PN WC200122920-A2.

XX PD 05-APR-2001.

XX PF 28-SEP-2000; 2000WO-US26524.

XX PR 29-SEP-1999; 99US-0157137.

XX PR 03-NOV-1999; 99US-0163280.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX DR WPI; 2001-235357/24.

XX DR N-PSDB; AAH35019.

XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers - Claim 11; Page 7851-7854; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patients own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent sequences used in the exemplification of the present invention.
 N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for

CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 869 AA;

alignment_scores:

Quality: 427.00 Length: 82
Ratio: 5.207 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-14 x AAG75614

Align seg 1/1 to: AAG75614 from: 1 to: 869

3 CTGAAGCGGGAATTCACGGGGCAGTCTCATTAATCTGACTTGGACAGC 52
|||||
714 LeulysAlaGluIleHisGlySerLeuIleAsnLeuThrTrpThrAl 730
53 TCCTGGGGATGATTATGACCATGGAACAGCTCACAAGTATATCATTCGAA 102
|||||
730 aProGlyAspAspTyrAspHisGlyThrAlaHisLysTyrIleIleArgI 747
103 TAAGTACAAGTATTCTTGATCTCAGACAGCAAGTTCAATGAATCTCTCAA 152
|||||
747 leSerThrSerIleLeuAspLeuArgAspLysPheAsnGluSerLeuGln 763
153 GTGAATACTACTGCTCTCATCCCAAGGAGCCAACTCTGAGGAAGTCTT 202
|||||
764 ValAsnThrThrAlaLeuIleProLysGluAlaAsnSerGluGluValPh 780
203 TTTGTTTAAACCAAGAACATTACTTTTGAATGGCAGACAGATCTT 248
|||||
780 eLeuPheLysProGluAsnIleThrPheGluAsnGlyThrAspLeu 795

seq_name: /SIDS2/gcgdata/geneseq/AA1999.DAT: AAB74824

seq_documentation_block:

ID AAB74824 standard; Protein: 914 AA.

AC AAB74824;

DT 13-JUN-2001 (first entry)

DE Human ICACC-1 protein sequence.

XX ICACC-1; ICACC-2; asthma; atopic allergy; asthma-related disorder;
KW Interleukin 9 induced calcium activated chloride channel; IL-9;
KW calcium activated chloride channel; anti-allergic; anti-asthmatic;
KW anti-inflammatory; immunomodulatory; cystic fibrosis;
KW inflammatory bowel disease; autoimmune disease.

XX Homo sapiens.

OS WO9944620-A1.

PN 10-SEP-1999.

XX 03-MAR-1999; 99WO-US04703.

XX 03-MAR-1998; 98US-0076815.

PA (MAGA-) MAGAININ PHARM INC.

XX Holroyd KJ, Levitt RC, Maloy WL, Louahed J, McLane M;

PI Nicolaides NC, Zhou Y, Dong Q;

DR WPI; 1999-550979/46.

DR N-PSDB; AAF81927.

XX New nucleic acid encoding calcium activated chloride channel, used to

PT identify, e.g. specific modulators for treating atopic allergy -

XX

PS Claim 11; Fig 4B; 75pp; English.

XX The present sequence represents the human interleukin 9 (IL-9) induced
CC calcium activated chloride channel 1 (ICACC-1) protein. ICACC proteins
CC have anti-allergic, anti-asthmatic, anti-inflammatory and
CC immunomodulatory activities. Compounds (A) that downregulate ICACC are
CC used to alleviate asthma (or more generally atopic allergy), while those
CC (B) that upregulate ICACC are used to treat cystic fibrosis (CF) and
CC inflammatory bowel disease (IBD) (or other autoimmune diseases).
CC Measuring levels of ICACC-1 can be used in the diagnosis of asthma
CC (increased levels) or IBD (reduced levels), also for monitoring
CC treatment of these conditions. The ICACC proteins can be used:
CC (i) to raise specific antibodies (Ab), useful: (a) as immunosay
CC reagents, and (b) as therapeutic (A); (ii) as ligands or mutants, as
CC (A) to disrupt bonding between ICACC and its ligands; and (iii) to
CC identify modulators and binding partners. ICACC polynucleotides can be
CC used to generate transgenic animals or recombinant cells, used to screen
CC for antagonists, also as a source of therapeutic antisense agents or
CC diagnostic probes (for quantifying mRNA expression, e.g. for
CC identification of modulators).

XX Sequence 914 AA;

alignment_scores:

Quality: 427.00 Length: 82
Ratio: 5.207 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-14 x AAB74824

Align seg 1/1 to: AAB74824 from: 1 to: 914

13 CTGAAGCGGGAATTCACGGGGCAGTCTCATTAATCTGACTTGGACAGC 52
|||||
752 LeulysAlaGluIleHisGlySerLeuIleAsnLeuThrTrpThrAl 775
53 TCCTGGGGATGATTATGACCATGGAACAGCTCACAAGTATATCATTCGAA 102
|||||
775 aProGlyAspAspTyrAspHisGlyThrAlaHisLysTyrIleIleArgI 792
103 TAAGTACAAGTATTCTTGATCTCAGACAGCAAGTTCAATGAATCTCTCAA 152
|||||
792 leSerThrSerIleLeuAspLeuArgAspLysPheAsnGluSerLeuGln 808
153 GTGAATACTACTGCTCTCATCCCAAGGAGCCAACTCTGAGGAAGTCTT 202
|||||
802 ValAsnThrThrAlaLeuIleProLysGluAlaAsnSerGluGluValPh 825
203 TTTGTTTAAACCAAGAACATTACTTTTGAATGGCAGACAGATCTT 248
|||||
825 eLeuPheLysProGluAsnIleThrPheGluAsnGlyThrAspLeu 840

seq_name: /SIDS2/gcgdata/geneseq/AA2001.DAT: AAM24514

seq_documentation_block:

ID AAM24514 standard; Protein: 914 AA.

AC AAM24514;

DT 12-OCT-2001 (first entry)

DE C902P predicted amino acid sequence.

XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;

XX immunogenic; gene therapy; vaccine; colonic cancer.

OS Homo sapiens.

XX WO200149716-A2.

XX 12-JUL-2001.

XX 29-DEC-2000; 2000WO-US5596.
XX 30-DEC-1999; 99US-0476296.
XX 10-JAN-2000; 2000US-0480321.
XX 15-FEB-2000; 2000US-0504629.
XX 06-MAR-2000; 2000US-0519444.
XX 19-MAY-2000; 2000US-0575251.
XX 29-JUN-2000; 2000US-0609448.
XX 28-AUG-2000; 2000US-0649811.
XX (CORI-) CORIXA CORP.
XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
XX King GE, Wang T, Jiang Y;
XX WPI; 2001-441847/47.
XX Colon tumor associated proteins and nucleic acids useful for the
XX prevention, diagnosis and treatment of colonic cancer -
XX
XX Claim 2; Page 440-443; 472pp; English.
XX
XX The present invention describes colon tumour associated proteins (I) and
XX the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX (II) may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate colon tumour associated protein (TCAP)
XX expression, such as colonic cancer. For example, (I) and (II) may be
XX used to treat disorders associated with decreased expression by
XX rectifying mutations or deletions in a patient's genome that affect the
XX activity of TCAPs by expressing inactive proteins or to supplement the
XX patients own production of them. Additionally, (II) may be used to
XX produce the TCAP proteins, by inserting the nucleic acids into a host
XX cell culturing the cell to express the protein. (II) and its
XX complementary sequences may also be used as DNA probes in diagnostic
XX polymerase chain reaction (PCR) and hybridisation assays to detect and
XX quantitate the presence of similar nucleic acids in samples, and
XX therefore which patients may be in need of restorative therapy. (I) may
XX also be used as antigens in the production of antibodies against TCAPs
XX and in assays to identify modulators of TCAP expression and activity.
XX Anti-(I) antibodies and antagonists may also be used to down regulate
XX TCAP expression and activity. The anti-(I) antibodies may also be used
XX as diagnostic agents for detecting the presence of TCAPs in samples
XX (e.g. by enzyme linked immunosorbent assay (ELISA)). AA128460 to AA129512
XX and AA24494 to AA24523 represent nucleotide and amino acid sequences
XX given in the exemplification of the present invention.
XX
XX Sequence 914 AA;
XX
XX alignment_scores:
XX Quality: 427.00 Length: 82
XX Ratio: 5.207 Gaps: 0
XX Percent Similarity: 100.000 Percent Identity: 100.000
XX
XX alignment_block:
XX US-09-049-696-14 x AA24514
XX
XX Align seg 1/1 to: AA24514 from: 1 to: 914
XX
XX 3 CTGAAGCGGAAATTCACGGGCGAGTCATTAATGACTGGACAGC 52
XX
XX 759 LeuLysAlaGluIleHisGlyGlySerLeuIleAsnLeuThrTrpAla 775
XX
XX 53 TCCTGGGATCATATGACCATGGAACAGCTCACAAAGTATATCATTCGAA 102
XX
XX 775 aProGlyAspAspTyrAspHisGlyThrAlaHisLysTyrIleIleArgI 792
XX
XX 103 TAAGTACAAGTATCTTGATCTCAGACACAAAGTTCATGATGATCTCTCAA 152
XX
XX 792 leSerThrSerIleLeuAspLeuArgAspLysPheAsnGluSerLeuGln 808

153 GTGAATACTACTGCTCTCATCCCAAGAGCCAACTCTCGAGGAGTCTT 202
|||||
809 ValAsnThrThrAlaLeuIleProLysGluAlaAsnSerGluGluValPh 825
203 TTTGTTTAAACACAGAAACATTACTTTTGAATAATGGCACAGATCTT 248
|||||
825 eleuPhelysProGluAsnIleThrPheGluAsnGlyThrAspLeu 840
seq name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.AAB73716
seq documentation_block:
ID AAB73716 standard; Protein; 914 AA.
XX
XX AAB73716;
XX
XX 11-SEP-2001 (first entry)
XX
XX Human CLCAL protein, SEQ ID NO:2.
XX
XX Human CLCAL; goblet cell; mouse Gob-5 orthologue; drug screening;
XX expression inhibition; antisense therapy; gene therapy;
XX chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Misc-difference 152 /note= "Encoded by AGG in AAH46124"
XX
XX WO200138530-A1.
XX
XX 31-MAY-2001.
XX
XX 22-NOV-2000; 2000WO-JP08232.
XX
XX 24-NOV-1999; 99JP-0333479.
XX 27-APR-2000; 2000JP-0127589.
XX
XX (TAKE) TAKEDA CHEM IND LTD.
XX
XX Nakanishi A, Morita S;
XX
XX WFI; 2001-355935/37.
XX N-PSDB; AAH46102, AAH46124.
XX
XX New antisense nucleotide, useful for treatment and prevention of
XX bronchial asthma and chronic obstructive pulmonary disease -
XX
XX Claim 2; Page 76-80; 104pp; Japanese.
XX
XX The invention relates to an antisense nucleotide targetted to the mouse
XX Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,
XX the CLCAL gene (coding sequence shown in AAH46102). The invention also
XX relates to an antibody specific for the Gob-5 protein, medical and
XX diagnostic compositions containing the antisense nucleotide or the
XX antibody, and methods and kits for screening for compounds which inhibit
XX the protein. Gob-5 and CLCAL are proteins expressed by goblet cells.
XX The antisense oligonucleotides and antibody are therefore useful for the
XX treatment and prevention of bronchial asthma and chronic obstructive
XX pulmonary disease. The present sequence represents human CLCAL protein.
XX
XX Sequence 914 AA;
XX
XX alignment_scores:
XX Quality: 427.00 Length: 82
XX Ratio: 5.207 Gaps: 0
XX Percent Similarity: 100.000 Percent Identity: 100.000
XX
XX alignment_block:
XX US-09-049-696-14 x AAB73716
XX
XX Align seg 1/1 to: AAB73716 from: 1 to: 914

3 CTGAAGCGGGAATTCACGGGGCAGTCTCATTAATCTGACTTGGACAGC 52
|||||
759 LeuLysAlaGluIleHisGlySerLeuIleAsnLeuThrTrpThrAl 775
53 TCCTGGGAGTATATCACCATGGAACAGCTCACAAGTATATATTCGAA 102
|||||
775 aProGlyAspPyrAspHisGlyThrAlaHisLysTyrIleIleArgI 792
103 TAAGTACAAGTATCTTGATCTCAGAGACAAGTTCATGAATCTCTTCAA 152
|||||
792 leSerThrSerIleLeuAspLeuArgAspLysPheAsnGluSerLeuGln 808
153 GTGAATACTACTGCTCTCATCCCAAGGAGCCAACTCTGAGGAAGTCTT 202
809 ValAsnThrThrAlaLeuIleProLysGluAlaAsnSerGluGluValPh 825
203 TTTGTTTTAAACCGAAGAACTTACTTTTGAAATGGCAGACATCTT 248
825 eLeuPheLysProGluAsnIleThrPheGluAsnGlyThrAspLeu 840

seq_name: /SIDS2/gcgdata/geneseq/AA2001.DAT: AAB74733

seq_documentation_block:
ID AAB74733 standard; Protein; 914 AA.

AC AAB74733;

12-JUN-2001 (first entry)

Human secreted protein sequence encoded by gene 1 SEQ ID NO:42.

Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
dermatological; immunosuppressive; antiinflammatory; anti-HIV;
immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;
antialzheimer; antiparkinsonian; antimicrobial; vulnery; gene therapy;
immune disorder; hyperproliferative disorder; cardiovascular disease;
cancer; angiogenic disorder; neurological disorder; infectious disease;
wound healing; regeneration; chemotaxis; chromosome 1.

Homo sapiens.

WO200112775-A2.

22-FEB-2001.

16-AUG-2000; 2000WO-US22325.

17-AUG-1999; 99US-0149182.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ni J, Florence KA, Fiscella M, Wei P, Baker KP;
Birse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;

WPI; 2001-147550/15.

N-PSDB; AAF81787.

Nucleic acids encoding 25 human secreted polypeptides, useful for
preventing, diagnosing and/or treating e.g. cancers, Parkinson's
disease and diabetic retinopathy -

Claim 11; Page 459-460; 485pp; English.

AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733
to AAB74772. Human secreted proteins can have activities based on the
tissues and cells they are expressed in. Example of activities include:
immunomodulatory; antisclerotic; dermatological; immunosuppressive;
antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;
vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic;
anticonvulsant; antialzheimer; antiparkinsonian; antimicrobial; and
vulnery. Human secreted proteins can be used in gene therapy and

CC vaccine. Human secreted protein nucleotide sequences (NAM1) and proteins
CC (PEP1) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. For example, NAM1
CC and PEP1 may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patients genome
CC that affect the activity of proteins by expressing inactive proteins or
CC to supplement the patients own production of polypeptides. Disorders that
CC may be prevented, diagnosed and/or treated include immune disorders,
CC hyperproliferative disorders (e.g. cancers), cardiovascular diseases,
CC angiogenic disorders, neurological disorders, infectious diseases and/or
CC for promoting wound healing, regeneration and/or chemotaxis. AAF81778 to
CC AAF81786 and AAB74732 represent sequences used in the exemplification of
CC the present invention.

XX
SQ Sequence 914 AA;

alignment_scores:

Quality: 427.00 Length: 82
Ratio: 5.207 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-14 x AAB74733 ..

Align seg 1/1 to: AAB74733 from: 1 to: 914

13 CTGAAGCGGGAATTCACGGGGCAGTCTCATTAATCTGACTTGGACAGC 52
|||||
759 LeuLysAlaGluIleHisGlySerLeuIleAsnLeuThrTrpThrAl 775
53 TCCTGGGAGTATATCACCATGGAACAGCTCACAAGTATATATTCGAA 102
|||||
775 aProGlyAspPyrAspHisGlyThrAlaHisLysTyrIleIleArgI 792
103 TAAGTACAAGTATCTTGATCTCAGAGACAAGTTCATGAATCTCTTCAA 152
|||||
792 leSerThrSerIleLeuAspLeuArgAspLysPheAsnGluSerLeuGln 808
153 GTGAATACTACTGCTCTCATCCCAAGGAGCCAACTCTGAGGAAGTCTT 202
|||||
809 ValAsnThrThrAlaLeuIleProLysGluAlaAsnSerGluGluValPh 825
203 TTTGTTTTAAACCGAAGAACTTACTTTTGAAATGGCAGACATCTT 248
825 eLeuPheLysProGluAsnIleThrPheGluAsnGlyThrAspLeu 840

seq_name: /SIDS2/gcgdata/geneseq/AA2001.DAT: AAG75474

seq_documentation_block:

ID AAG75474 standard; Protein; 925 AA.

AC AAG75474;

ET 03-SEP-2001 (first entry)

Human colon cancer antigen protein SEQ ID NO:6238.

Human; colon cancer; colon cancer antigen; diagnosis; detection;
colorectal carcinoma; chromosome 1.

Homo sapiens.

WO200122920-A2.

05-APR-2001.

26-SEP-2000; 2000WO-US26524.

29-SEP-1999; 99US-0157137.

03-NOV-1999; 99US-0163280.

(HUMA-) HUMAN GENOME SCI INC.

```
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX KW WPI; 2001-235357/24.
XX DR N-PSDB; AAH34879.
XX XX
XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX PS Claim 11; Page 7686-7690; 9803pp; English.
XX
XX CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAB77789 represent sequences used in the exemplification of the
XX present invention.
XX CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX SQ Sequence 925 AA;

alignment_scores:
    Quality: 427.00      Length: 82
    Ratio: 5.207        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-09-049-696-14 x AAG75474 ..
Align seg 1/1 to: AAG75474 from: 1 to: 925

3 CTGAAGCGGAAATTCACGGGGCAGTCTCATTAATCTGACTTGGACAGC 52
|||||
770 LeuLysAlaGluIleHisGlySerLeuIleAsnLeuThrTrpThrAl 786

53 TCCTGGGATGATTATCACCAGCTCACAAGCTCACAAGTATATCATTCGAA 102
|||||
786 aProGlyAspAspTyrAspHisGlyThrAlaHisLysTyrIleIleArgI 803

103 TAAGTACAAGTATCTTGATCTCAGAGACAAAGTTCATTAATCTCTTCAA 152
|||||
803 leSerThrSerIleLeuAspLeuArgAspLysPheAsnGluSerLeuGln 819

153 GTGAATACTACTGCTCTCATCCCAAGGAGCAACTCTGAGGAAGTCTT 202
|||||
820 ValAsnThrThrAlaLeuIleProLysGluAlaAsnSerGluGluValph 836

203 TTGTGTTTAAACCAACAACTTACTTTTGAATAATGGCACAGATCTT 248
|||||
836 eLeuPheLysProGluAsnIleThrPheGluAsnGlyThrAspLeu 851

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT: AAG73854

seq_documentation_block:
ID AAG73854 standard; Protein; 552 AA.
XX
XX AC AAG73854;
XX
XX DT 03-SEP-2001 (first entry)
XX
```

```
DE Human colon cancer antigen protein SEQ ID NO:4618.
XX
XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; chromosome 1.
XX
XX OS Homo sapiens.
XX
XX PN WC200122920-A2.
XX
XX PD 05-APR-2001.
XX
XX PF 28-SEP-2000; 2000WO-US26524.
XX
XX PR 29-SEP-1999; 99US-0157137.
XX
XX PR 03-NOV-1999; 99US-0163280.
XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX DR N-PSDB; AAH33285.
XX
XX XX
XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX PS Claim 11; Page 6416-6419; 9803pp; English.
XX
XX CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAB77789 represent sequences used in the exemplification of the
XX present invention.
XX CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX SQ Sequence 552 AA;

alignment_scores:
    Quality: 421.00      Length: 82
    Ratio: 5.134        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 98.780

alignment_block:
US-09-049-696-14 x AAG73854 ..
Align seg 1/1 to: AAG73854 from: 1 to: 552

3 CTGAAGCGGAAATTCACGGGGCAGTCTCATTAATCTGACTTGGACAGC 52
|||||
397. LeuLysAlaGluIleHisGlySerLeuIleAsnLeuThrTrpThrAl 413

53 TCCTGGGATGATTATGACCAGTGGAAACAGCTCACAAGTATATCATTCGAA 102
|||||
413 aProGlyAspAspTyrAspHisGlyThrAlaHisLysTyrIleIleArgI 430

103 TAAGTACAAGTATCTTGATCTCAGAGACAAAGTTCATTAATCTCTTCAA 152
|||||
430 leSerThrSerIleLeuAspLeuArgAspLysPheAsnGluSerLeuGln 446
```


PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
XX
XX WPI: 2001-183260/18.
DR N-PSDB; AAF92092.
XX
XX Eighty four nucleic acids encoding PRO polypeptides, useful in
XX molecular biology, including use as hybridization probes, and in
XX chromosome and gene mapping.
XX
XX Claim 12; Fig 70; 278pp; English.
XX
XX The present sequence is a human PRO polypeptide (secreted and
XX transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
XX anti-PRO antibodies are useful for preparation of a medicament useful in
XX the treatment of a condition which is responsive to the PRO protein,
XX agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
XX employed as molecular weight markers for protein electrophoresis. The PRO
XX coding sequence has applications in molecular biology, including use as
XX hybridisation probes, and in chromosome and gene mapping.
XX
SQ Sequence 919 AA;

alignment_scores:
Quality: 259.50 Length: 82
Ratio: 3.761 Gaps: 1
Percent Similarity: 84.146 Percent Identity: 63.415

alignment_block:
US-09-049-696-14 x AAB87560 ..
Align seg 1/1 to: AAB87560 from: 1 to: 919

3 CTGAAGCGGGAATTCACGGGGCAGCTCATTAACTGACTTGGACAGC 52
761 LeuAspAlaThrValHisGluAspLysIleile...LeuThrTrpThrAl 776

53 TCCTGGGATGATTATGACCATGACGAGCTCAGAGTCAAGTATATATTCGAA 102
776 aproGlyAspAsnPheAspValGlyLysValGlnArgTyrIleileargi 793

103 TAAGTACAGTATTCTTGTGATCTCAGACAGCAAGTTCAATGAATCTCTCAA 152
793 leSerAlaSerIleLeuAspLeuArgAspSerPheAspAspAlaLeuGln 809

153 GTGAATACTGCTCTCATCCCAAGGAGCCCAACTCTGAGGAAGTCTT 202
810 ValAsnThrThrAspLeuSerProLysGluAlaAsnSerLysGluSerPh 826

203 TTGTGTTTAAACCAAGAAACATTTACTTTTGAATGCGACAGATCTT 248
826 eAlaPheLysProGluAsnIleSerGluGluAsnAlaThrHisIle 841

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.AAB65272
seq_documentation_block:
ID_AAB65272 standard; Protein: 919 AA.
XX
XX AAB65272;
XX
XX 02-APR-2001 (first entry)
XX
XX Human PRO1124 (UNQ562) protein sequence SEQ ID NO:379.
XX
XX Human; secreted and transmembrane protein; PRO: cytostatic;
KW cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
XX diagnostic assay.
XX
XX Homo sapiens.
XX
XX WO200073454-A1.
XX

PD 07-DEC-2000.
XX
XX 3Q-MAR-2000; 2000WO-US08439.
XX
XX 02-JUN-1999; 99WO-US12252.
XX 23-JUN-1999; 99US-0141037.
XX 07-JUL-1999; 99US-0143048.
XX 20-JUL-1999; 99US-0144758.
XX 26-JUL-1999; 99US-0145698.
XX 28-JUL-1999; 99US-0146222.
XX 17-AUG-1999; 99US-0149396.
XX 15-SEP-1999; 99WO-US21090.
XX 15-SEP-1999; 99WO-US21547.
XX 08-OCT-1999; 99US-0158663.
XX 30-NOV-1999; 99WO-US28313.
XX 01-DEC-1999; 99WO-US28301.
XX 16-DEC-1999; 99WO-US30095.
XX 20-DEC-1999; 99WO-US30911.
XX 05-JAN-2000; 2000WO-US00219.
XX 06-JAN-2000; 2000WO-US00376.
XX 11-FEB-2000; 2000WO-US03565.
XX 18-FEB-2000; 2000WO-US04341.
XX 22-FEB-2000; 2000WO-US04414.
XX 24-FEB-2000; 2000WO-US04914.
XX 24-FEB-2000; 2000WO-US05004.
XX 02-MAR-2000; 2000WO-US05841.
XX 15-MAR-2000; 2000WO-US06884.
XX 20-MAR-2000; 2000WO-US07377.
XX
XX (GETH) GENENTECH INC.
XX
XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
XX Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
XX Grimaldi CJ, Gurney AL, Kijavlin IJ, Napier MA, Pan J, Paoni NF;
XX Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
XX Zhang Z;
XX
XX WPI: 2001-032160/04.
XX N-PSDB; AAF44241.
XX
XX PRO polynucleotides used to produce polypeptides used to target
XX bioactive molecules such as toxins, radiolabels or antibodies, to
XX specific cells, to cause targeted cell death -
XX
XX Claim 12; Fig 274; 935pp; English.
XX
XX The present invention describes human secreted and transmembrane PRO
XX proteins. The PRO proteins have cytostatic activity. The PRO proteins
XX can be used for targeted delivery of bioactive molecules, such as
XX toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
XX sequences, and their fragments, can be used as hybridisation probes, in
XX chromosomal and gene mapping, and in the generation of anti-sense RNA
XX and DNA. They may also be used to produce transgenic animals which are
XX used to develop and screen therapeutically useful reagents. The PRO
XX nucleotide and protein sequence can be used for tissue typing and in
XX treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
XX AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
XX in the isolation of human PRO sequences. AAF44087 to AAF44269 and
XX AAB65154 to AAB65300 represent human PRO polynucleotide and protein
XX sequences given in the exemplification of the present invention.
XX
XX Sequence 919 AA;
XX

alignment_scores:
Quality: 259.50 Length: 82
Ratio: 3.761 Gaps: 1
Percent Similarity: 84.146 Percent Identity: 63.415

alignment_block:
US-09-049-696-14 x AAB65272 ..
Align seg 1/1 to: AAB65272 from: 1 to: 919

```

alignment_scores:..
  Quality: 159.00
  Ratio: 2.944
  Percent Similarity: 76.056
  Percent Identity: 42.254
alignment_block:
  OS-091049-696-14 x AAB45904
  Align seg 1/1 to: AAB45904 from: 1 to: 742
  33 ATTAATCTGACTTGGACAGCTCTGGGCGATGATTATGACCATGGACAGC 82
  580 LeuThrLeuSerTrpThrAlaProGlyGluAspPheAspGlnGlyGlnAl 136
  83 TCACAAGTATATCATTCGAAATAAGTACAAAGATTCCTGATCTCAGAGACA 132
  596 aThrSerTyTyrGluIleArgMetSerLysSerLeuGlnAsnIleGlnAspA 613
  133 AGTTCATGAATCTCTTCAAGTGAATACTACTGCTCTCATCCCAAGGAA 182
  613 spPheAsnAsnAlaIleLeuValAsnThrSerLysArgAsnProGlnGln 639
  183 GCCAACTCTCAGGGAAGTCCTTGTGTTTAAACCCAGAAACATTACTTTTCA 232
  640 AlaGlyIleArgGluIlePheThrPheSerProGlnIleSerThr..... 644
  233 AAATGGCACAGAT 245
  645 AsnGlyProGlu 648

```

133 AGTTCATGAATCTCTCAAGTAGTAATACTACTGCTCATCCAAAGGAA
613 spPheAsnAlaIleuValAsnThrSerLysArgAsnProGluGln
183 GCCAATCTCAGGAGAACGTCTTTTGTAAACACAGAAAAATATTCTTTGA
630 AlaGlyIleargGluIlePheThrPheSerProGluIleSerThr.... 644
233 AAATGGCACAGAT 245
645 AsnGlyProGlu 648

[illegible]

233 AAATGGCACAGAT 245
||||| :
645 AsnGlyProGlu 648

255 AAATGGCAGAT 243
||||| ::
645 AsnGlyproGlu 648

[illegible]

•

100

1000

This Page Blank (uspto)

OM of: US-09-049-696-13 to: SPTREMBL_17.* out_format : pfs

Date: Mar 30, 2002 2:47 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+ntp_model -DEV=xlp
-Q/cgnt2.1/USPTO_spool/US09049696/runat_28032002_145238_2053/app_query.fasta_1.12579
-DB=SPTREMBL_17 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-EGAPOP=6.000 -EGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09049696 @CGN1_1_705 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-049-696-13
Query length: 227
Database: SPTREMBL_17.*
Database sequences: 473505
Database length: 14627329
Search time (sec): 805.760000
```

score_list:

Sequence	Strd	Orig	Zscore	EScore	Len	Document
sp_human:095151	+	403.00	864.12	8.1e-41	914	1095151 homo sapiens (human). ca
sp_human:09UPC6	+	403.00	864.12	8.1e-41	914	109UPC6 homo sapiens (human). ca
sp_human:09UNF6	+	403.00	864.12	8.1e-41	914	109UNF6 homo sapiens (human). ca
sp_mammal:09TUB5	+	311.00	667.68	1.5e-29	917	109TUB5 mus scrofa (pig). epithe
sp_rodent:088826	+	227.50	477.99	2.6e-19	913	1088826 mus musculus (mouse). gc
sp_rodent:09D726	+	222.50	466.99	1.1e-18	913	109D726 mus musculus (mouse). ch
sp_human:09NXP1	+	209.50	443.39	4.3e-17	917	109NXP1 homo sapiens (human). ca
sp_human:09UNF7	+	209.50	438.35	4.2e-17	917	109UNF7 homo sapiens (human). ca
sp_mammal:018742	+	144.00	295.32	4.5e-09	820	1018742 bos taurus (bovine). lu
sp_mammal:018743	+	144.00	295.32	4.5e-09	820	1018743 bos taurus (bovine). lu
sp_mammal:018741	+	144.00	294.34	4.5e-09	905	1018741 bos taurus (bovine). lu
sp_rodent:09BQR4	+	136.50	277.86	3.7e-08	902	109BQR4 mus musculus (mouse). en
sp_rodent:088860	+	135.50	275.67	4.9e-08	901	1088860 mus musculus (mouse). ch
sp_rodent:09QX15	+	135.50	275.66	4.9e-08	902	109QX15 mus musculus (mouse). ca
sp_rodent:09QX17	+	135.50	264.66	2.0e-07	902	109QX17 mus musculus (mouse). ca
sp_human:09UQC9	+	123.50	248.92	1.4e-06	943	109UQC9 homo sapiens (human). ch
sp_human:09Y6N2	+	123.50	248.92	1.4e-06	943	109Y6N2 homo sapiens (human). ca
sp_invertebrate:09VE32	-	72.00	143.16	3.08	345	109VE32 drosophila melanogaste
sp_bacteria:085154	-	69.00	133.31	7.07	532	1085154 photorhabdus luminescens
sp_bacteria:095822	-	68.00	139.85	9.79	166	1095822 synechococcus sp. (strai
sp_invertebrate:09BLJ1	-	67.00	124.66	12.17	937	109BLJ1 ciona intestinalis. ci
sp_bacteria:071093	+	66.50	127.34	14.27	566	1071093 bacteroides thetaiotaom
sp_invertebrate:09VAU0	-	66.00	130.32	16.77	329	109VAU0 drosophila melanogaste
sp_invertebrate:09N8C8	-	66.00	129.70	16.72	357	109N8C8 drosophila melanogaste
sp_plant:09M8U0	+	66.00	127.67	16.55	468	109M8U0 arabidopsis thaliana (mc
sp_plant:064397	+	66.00	125.30	16.36	642	1064397 lycopersicon esculentum
sp_human:075189	+	65.50	120.05	18.46	1115	1075189 homo sapiens (human). k
sp_plant:09M1R2	+	65.00	124.54	21.85	530	109M1R2 arabidopsis thaliana (mc
sp_rodent:09QW15	-	65.00	122.67	21.65	680	109QW15 mus sp. beta 3 integrin,
sp_rodent:09QW16	-	65.00	122.21	21.60	723	109QW16 rattus sp. beta 3 integr
sp_bacteria:069204	+	64.50	133.22	26.41	144	1069204 actinosynnema pretiosum
sp_plant:09ZV03	+	64.50	127.08	25.62	326	109ZV03 arabidopsis thaliana (mc
sp_bacteria:007426	+	64.50	126.13	25.50	370	1007426 mycobacterium tuberculo
sp_invertebrate:09N9W3	-	64.50	118.78	24.59	986	109N9W3 haemulonchus contortus.
sp_invertebrate:09GND1	-	64.00	128.77	29.31	225	109GND1 leishmania major. 1928
sp_plant:09AUB6	+	64.00	127.43	29.72	269	109AUB6 lycopersicon pennellii
sp_plant:09FLY7	+	64.00	121.84	28.90	566	109FLY7 arabidopsis thaliana (mc
sp_plant:09FLY8	+	64.00	121.83	28.90	567	109FLY8 arabidopsis thaliana (mc
sp_mammal:09TUN5	-	64.00	119.40	28.56	784	109TUN5 mus scrofa (pig). glycop
sp_invertebrate:09XV10	+	63.50	127.09	34.35	243	109XV10 caenorhabditis elegans

```
sp_virus:09JF66 + 63.50 124.86 33.97 327 109JF66 vaccinia virus (strai
sp_plant:09ART1 + 63.50 122.11 33.51 472 109ART1 oryza sativa (rice).
sp_human:09H0B7 + 63.50 118.24 32.87 790 109H0B7 homo sapiens (human).
sp_virus:057235 + 63.00 125.23 39.40 269 1057235 vaccinia virus (strai
sp_bacteria:09X8X9 + 63.00 122.14 38.80 406 109X8X9 streptomyces coelicol
```

seq_name: sp_human:095151

seq_documentation_block:

```
ID 095151 PRELIMINARY; PRT; 914 AA.
AC 095151;
DT 01-MAY-1999 (TREMREL. 10, Created)
DT 01-MAY-1999 (TREMREL. 10, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Elble R.C., Ji H.L., Schreier K.D., Fuller C.M.,
RA Pauli B.O.;
RT "Genomic cloning, molecular characterization, and functional analysis
RT of human CLCAL, the first human member of the family of Ca2+-activated
RL Genomics 54:200-214(1998).
DR EMBL: AF039400; AAC95428.1;
DR InterPro: IPR000131; ATPase_gamma.
DR IctPro: IPR002035; VWFA.
DR PROSITE: P500153; ATPASE_GAMMA; UNKNOWN_1.
DR PROSITE: P500234; VWFA; 1.
DR SMART: SM00327; VWFA; 1.
SQ SEQUENCE 914 AA, 100174 MW, 7066440C46557FA3 CRC64;
```

```
alignment_scores:
Quality: 403.00 Length: 75
Ratio: 5.373 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment_block:

```
US-09-049-696-13 x 095151
Align seg 1/1 to: 095151 from: 1 to: 914
1 GTGCGGCTCTGGGAGGAGTTAAGCAGCAGCGAGAGTGATACCCCA 50
|||||
675 ValArgAlaLeuGlyValAsnAlaAlaArgArgValIleProGl 691
51 GCAGATGGAGCAGCTACATACCTGGCTGGATTGAGATGAGAAATAC 100
|||||
691 nGlnSerGlyAlaLeuTyrIleProGlyTyrPilleGluAsnAspGluLeG 708
101 AATGAATCCACCAAGACCTGAAATTAATAGATGATGTTCAACACAAG 150
708 InTrpAsnProProArgProGluIleAsnLysAspValGlnHisLys 724
151 CAAGTGTGTTTACAGACATCTCCGGAGGCTCATTTGTGGCTTCGA 200
|||||
725 GlnValCysPheSerArgThrSerSerGlyGlySerPheValAlaSerAs 741
201 TGTCCTCAATGCTCCCATCTGAT 225
741 pValProAsnAlaProIleProAsp 749
```

seq_name: sp_human:09UPC6

seq_documentation_block:

```
ID 09UPC6 PRELIMINARY; PRT; 914 AA.
```

AC Q9UPC6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CALCIUM-DEPENDENT CHLORIDE CHANNEL-1.
GN HCLCAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99047526; PubMed=9828122;
RA Gruber A.D., Elble R.C., Ji H.L., Schreier K.D., Fuller C.M.,
RA Pauli B.U.;
RT "Genomic cloning, molecular characterization, and functional analysis
RT of human CLCAL, the first human member of the family of Ca2+-activated
RT Cl-channel proteins.";
RL Genomics 54:200-214(1998).
RL EMBL; AF039401; AAC95429.1; -;
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPase_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWFA; 1.
SQ SEQUENCE 914 AA; 100160 MW; 7761362B66752AA3 CRC64;

alignment_scores:
Quality: 403.00 Length: 75
Ratio: 5.373 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-13 x Q9UPC6 ..
Align seg 1/1 to: Q9UPC6 from: 1 to: 914

1 GTGCGGGCTCTGGAGAGTTAAACGACGAGGAGAGTATACCCCA 50
|||||
675 ValArgAlaLeuGlyValAsnAlaAlaArgArgValIleProG1 691

51 GCAGAGTGGAGCAGTGTACATACCTGGCTGGATGAGATGAAATAC 100
|||||
691 nGlnSerGlyAlaLeuTyrlleProGlyIleGluAsnAspGluileG 708

101 AATGAATCCACCAAGACCTGAAATTAAGATGATGTTCAACACAAG 150
|||||
708 InTrpAsnProProArgProGluIleAsnLysAspValGlnHisLys 724

151 CAAGTGTGTTTCAGCAGAACATCTCGGAGGCTCATTTGGCTTCTGA 200
|||||
725 GlnValCysPheSerArgThrSerGlyGlySerPheValAlaSerAs 741

201 TGTCCCAATGCTCCCATACCTGAT 225
741 pValProAsnAlaProIleProasp 749

seq_name: sp_human.Q9UNF6
seq_documentation_block:
ID Q9UNF6 PRELIMINARY; PRT; 914 AA.
AC Q9UNF6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CALCIUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 1.
GN CACCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE, AND COLON;
RX MEDLINE=99364503; PubMed=10437792;
RA Agnel M., Vermet T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
RT chloride channel (CaCC) family predominantly expressed in the
RT digestive tract and trachea.";
RL FEBS Lett. 455:295-301(1999).
DR EMBL; AF127036; AAD25487.1; -;
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS00153; ATPase_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 914 AA; 100202 MW; B58B451D97887D6B CRC64;

alignment_scores:
Quality: 403.00 Length: 75
Ratio: 5.373 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-049-696-13 x Q9UNF6 ..
Align seg 1/1 to: Q9UNF6 from: 1 to: 914

1 GTGCGGGCTCTGGAGAGTTAAACGACGAGGAGAGTATACCCCA 50
|||||
675 ValArgAlaLeuGlyValAsnAlaAlaArgArgValIleProG1 691

51 GCAGAGTGGAGCAGTGTACATACCTGGCTGGATGAGATGAAATAC 100
|||||
691 nGlnSerGlyAlaLeuTyrlleProGlyIleGluAsnAspGluileG 708

101 AATGAATCCACCAAGACCTGAAATTAAGATGATGTTCAACACAAG 150
|||||
708 InTrpAsnProProArgProGluIleAsnLysAspValGlnHisLys 724

151 CAAGTGTGTTTCAGCAGAACATCTCGGAGGCTCATTTGGCTTCTGA 200
|||||
725 GlnValCysPheSerArgThrSerGlyGlySerPheValAlaSerAs 741

201 TGTCCCAATGCTCCCATACCTGAT 225
741 pValProAsnAlaProIleProasp 749

seq_name: sp_mammal.Q9TUB5
seq_documentation_block:
ID Q9TUB5 PRELIMINARY; PRT; 917 AA.
AC Q9TUB5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN.
GN AECC.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
CX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ILEAL MUCOSA;
RA Gaspar K.J., Gabriel S.E., Racette K.J., Forsyth G.W.;
RT "A cDNA involved in porcine exocrine chloride conductance.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF095584; AAF00077.1; -;
DR InterPro; IPR000131; ATPase_gamma.
DR InterPro; IPR002035; VWFA.
DR PROSITE; PS00153; ATPase_GAMMA; UNKNOWN_1.
DR PROSITE; PS50234; VWFA; 1.
DR SMART; SM00327; VWFA; 1.

ATTN: SEC 1/1 CO: 000020 FROM: 1 CO: 313

.PERCENT SIMILARITY: 79.730 PERCENT IDENTITY: 60.811

alignment_block:

US-09-049-696-13 x Q9D726 ..

Align seg 1/1 to: Q9D726 from: 1 to: 913

```
7 GCTCTGGGAGGAGTTAACGCGAGGAGGAGTATACCCGAGAGAG 56
|||||
678 AlaLeuGlyGlyValThrSerArgGlnArgAlaAlaProPlysAs 694
|||||
57 TGGGAGCAGTGTACATCTGGCTGGATTCAGAAATGATGATGGA 106
|||||
694 nArgProMetTyrlleaspGlyTrpIleuaspGlyGluValArgMet 711
|||||
107 ATCCACCAAGACCTGAAATTAATAGGATGATGTTCAACACAGCAAGTC 156
|||||
711 snProProArgProGluThrSerTyr.....ValGlnAspLysGlnLeu 725
|||||
157 TGTTCAGCAGACATCCCTGGGAGGCTATTTGTGGCTTCTGATGCC 206
|||||
726 CysPheSerArgThrSerSerGlyGlySerPheValAlaThrAsnValPr 742
|||||
207 A...AATGCTCCCATACCTGAT 225
|
742 oAlaAlaAlaProIleProasp 749
|
```

seq_name: sp_human:Q9NXP1

seq_documentation_block:

ID Q9NXP1 PRELIMINARY; PRT; 469 AA.

```
AC Q9NXP1;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
DE CDNA FLJ20131 FTS, CLONE COL06357.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Ohayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000138; BAA90969.1; -.
SQ SEQUENCE 469 AA; 50694 MW; 6765986C886B3E56 CRC64;
```

alignment_scores:

Quality: 209.50 Length: 75
Ratio: 3.612 Gaps: 1
Percent Similarity: 77.333 Percent Identity: 56.000

alignment_block:

US-09-049-696-13 x Q9NXP1 ..

Align seg 1/1 to: Q9NXP1 from: 1 to: 469

```
1 GTGCGGCTCTGGAGGAGTTAACGCGAGGAGGAGTATACCCCA 50
|||||
230 ValArgAlaHisGlyGlyAlaAsnThrAlaArgLeuLysLeuArgPro 246
|||||
51 GCAGAGTGGAGCACTGTACATACCTGGCTGGATTGAGATGATAATAC 100
|||||
246 oLeuAsnArgAlaAlaTyrlleProGlyTrpValValAsnGlyGluIle 263
|||||
101 AATGGATCCACCAAGACCTGAAATTAATAGGATGATGTTCAACACAG 150
|||||
263 luAlaAsnProProArgProGluIleAsp...GluAspThrGlnThr 278
|||||
```

```
151 CNAAGTGTGTTTCAGCAGAACATCTCGGAGGCTCATTGTGGCTTCTGA 200
|||||
279 LeuGluAspPheSerArgThrAlaSerGlyAlaPheValValSerGI 295
|||||
201 TGTCCCAAAATGCTCCCATACCTGAT 225
|||||
295 nValProSerLeuProLeuProasp 303
|||||
seq_name: sp_human:Q9UNF7
```

```
seq_documentation_block:
ID Q9UNF7 PRELIMINARY; PRT; 917 AA.
AC Q9UNF7;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE CALCINUM-ACTIVATED CHLORIDE CHANNEL PROTEIN 2.
GN CACCC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Agnel M., Vermet T., Culouscou J.M.;
RT "Identification of three novel members of the calcium-dependent
RT chloride channel (CaCC) family predominantly expressed in the
RT digestive tract and trachea.";
RL FEBS Lett. 455:295-301(1999).
DR EMBL; AF127035; AAD48398.1; -.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00092; vwa; 1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
SQ SEQUENCE 917 AA; 101153 MW; EA01090D781BEB95 CRC64;
```

alignment_scores:

Quality: 209.50 Length: 75
Ratio: 3.612 Gaps: 1
Percent Similarity: 77.333 Percent Identity: 56.000

alignment_block:

US-09-049-696-13 x Q9UNF7 ..

Align seg 1/1 to: Q9UNF7 from: 1 to: 917

```
1 GTGCGGCTCTGGAGGAGTTAACGCGAGGAGGAGTATACCCCA 50
|||||
678 ValArgAlaHisGlyGlyAlaAsnThrAlaArgLeuLysLeuArgPro 694
|||||
51 GCAGAGTGGAGCACTGTACATACCTGGCTGGATTGAGATGATAATAC 100
|||||
694 oLeuAsnArgAlaAlaTyrlleProGlyTrpValValAsnGlyGluIle 711
|||||
101 AATGGATCCACCAAGACCTGAAATTAATAGGATGATGTTCAACACAG 150
|||||
711 luAlaAsnProProArgProGluIleAsp...GluAspThrGlnThr 726
|||||
151 CNAAGTGTGTTTCAGCAGAACATCTCGGAGGCTCATTGTGGCTTCTGA 200
|||||
727 LeuGluAspPheSerArgThrAlaSerGlyAlaPheValValSerGI 743
|||||
201 TGTCCCAAAATGCTCCCATACCTGAT 225
|||||
743 nValProSerLeuProLeuProasp 751
|||||
```

seq_name: sp_mammal:O18742

seq_documentation_block:

ID O18742 PRELIMINARY; PRT; 794 AA.

```

AC O18742;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001262; AAB86530.1;
DR InterPro; IPRO02035; VWFA.
DR PROSITE; PS0234; VWFA.
DR SMART; SM00327; VWA; 1.
DR SEQ 794 AA; 88509 MW; B695E7256FC2632 CRC64;

alignment_scores:
  Quality: 144.00      Length: 69
  Ratio: 2.880        Gaps: 0
Percent Similarity: 72.464 Percent Identity: 44.928

alignment_block:
US-09-049-696-13 x O18742 ..
Align seg 1/1 to: O18742 from: 1 to: 794

1 GTGGGGCTCGGAGGAGTAAAGCGAGCGAGAGGAGATGATACCCCA 50
||||:||||| :|||:||||| :|||:||||| :|||:|||||
683 ValHisAlaGlnAlaArgAsnThrAlaArgLeuAsnLeuArgGlnPr 699

51 GCAGAGTGGAGCACTGTACATACCTGGCTGGATTGAGAATGAAATAC 100
||||:||||| :|||:||||| :|||:||||| :|||:|||||
699 oGlnAsnLysValLeuTyrValProGlyTyrValGluAsnGlyLysIleI 716

101 AATGGAATCCACCAAGACCTGAAATTAAGGATGATGTTCAACACAAG 150
||||:||||| :|||:||||| :|||:||||| :|||:|||||
716 leLeuAsnProArgProGluValLysAspAspLeuAlaLysAlaLys 732

151 CAAGTGTGTTTCAGCAACATCTCCGGAGGCTCATTTGGCTTCTGA 200
||||:||||| :|||:||||| :|||:||||| :|||:|||||
733 IleGluAspPheSerArgLeuThrSerGlyGlySerPheThrValSerGI 749

201 TGTCCCA 207
749 yAlaPro 751

seq_name: sp_mammal:O18742

seq_documentation_block:
ID O18741 PRELIMINARY; PRT; 820 AA.
AC O18741;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001261; AAB86529.1;
DR InterPro; IPRO02035; VWFA.
DR InterPro; IPRO03961; FN_III.
DR Pfam; PF00041; fn3; 1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR SEQ 905 AA; 101005 MW; D4CA6E7919115E88 CRC64;

alignment_scores:
  Quality: 144.00      Length: 69
  Ratio: 2.880        Gaps: 0
Percent Similarity: 72.464 Percent Identity: 44.928

alignment_block:
US-09-049-696-13 x O18741 ..

```

```

DR EMBL; AF001263; AAB86531.1;
DR InterPro; IPRO02035; VWFA.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR SEQ 820 AA; 91464 MW; 4E01330AEF5A1494 CRC64;

alignment_scores:
  Quality: 144.00      Length: 69
  Ratio: 2.880        Gaps: 0
Percent Similarity: 72.464 Percent Identity: 44.928

alignment_block:
US-09-049-696-13 x O18743 ..
Align seg 1/1 to: O18743 from: 1 to: 820

1 GTGGGGCTCGGAGGAGTAAAGCGAGCGAGAGGAGATGATACCCCA 50
||||:||||| :|||:||||| :|||:||||| :|||:|||||
683 ValHisAlaGlnAlaArgAsnThrAlaArgLeuAsnLeuArgGlnPr 699

51 GCAGAGTGGAGCACTGTACATACCTGGCTGGATTGAGAATGAAATAC 100
||||:||||| :|||:||||| :|||:||||| :|||:|||||
699 oGlnAsnLysValLeuTyrValProGlyTyrValGluAsnGlyLysIleI 716

101 AATGGAATCCACCAAGACCTGAAATTAAGGATGATGTTCAACACAAG 150
||||:||||| :|||:||||| :|||:||||| :|||:|||||
716 leLeuAsnProArgProGluValLysAspAspLeuAlaLysAlaLys 732

151 CAAGTGTGTTTCAGCAACATCTCCGGAGGCTCATTTGGCTTCTGA 200
||||:||||| :|||:||||| :|||:||||| :|||:|||||
733 IleGluAspPheSerArgLeuThrSerGlyGlySerPheThrValSerGI 749

201 TGTCCCA 207
743 yAlaPro 751

seq_name: sp_mammal:O18741

seq_documentation_block:
ID O18741 PRELIMINARY; PRT; 905 AA.
AC O18741;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LU-ECAM-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Elble R.C., Widom J., Gruber A.D., Abdel-Ghany M., Levine R.,
RA Goodwin A., Pauli B.U.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001261; AAB86529.1;
DR InterPro; IPRO02035; VWFA.
DR InterPro; IPRO03961; FN_III.
DR Pfam; PF00041; fn3; 1.
DR PROSITE; PS0234; VWFA; 1.
DR SMART; SM00327; VWA; 1.
DR SEQ 905 AA; 101005 MW; D4CA6E7919115E88 CRC64;

alignment_scores:
  Quality: 144.00      Length: 69
  Ratio: 2.880        Gaps: 0
Percent Similarity: 72.464 Percent Identity: 44.928

alignment_block:
US-09-049-696-13 x O18741 ..

```

Align seg 1/1 to: 018741 from: 1 to: 905

```

1 GTGCGGCTCTGGAGAGTTAACGAGCAGCAGGAGAGTATACCCCA 50
|||||:||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||:
683 ValHisAlaGlnAlaArgAsnAsnThrAlaArgLeuAsnLeuArgGlnPr 699
51 GCAGAGTGAGGAGCACTGTACATACCTGGCTGGCTGGATGTGAGATGAAATAC 100
|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||:
699 oGlnAsnLysValLeuTyrValProGlyTyrValGluAsnGlyLysIleI 716
101 AATGGAATCCACAGACCTGAAATTAATAAGGATGATGTTCAACACAAG 150
|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||:
716 leuAsnProArgProGluValLysAspLeuAlaLysAlaLys 732
151 CAAGTGTGTTTTCAGCAAACTCTCGGAGGCTCATTTGTGGCTTCTGA 200
|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||:
733 IleGluAspPheSerArgLeuThrSerGlyGlySerPheThrValSerGl 749
201 TGTCCCA 207
749 yAlaPro 751

```

seq_name: sp_rodent:Q9EQR4

seq_documentation_block:

```

ID Q9EQR4 PRELIMINARY; PRT; 902 AA.
AC Q9EQR4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ENDOTHELIAL CHLORIDE CHANNEL.
GN CLCA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Eble R.C., Paul B.;
RT "A murine endothelial homolog of Lu-ECAM-1 associated with lung
  melanoma metastasis."
RL Submitted (DPC-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AF115852; AKA47626.1;
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR002035; VWFA.
DR SMART: SM00060; FN3; 1.
DR SMART: SM00327; VWA; 1.
DR PROSITE: PS50234; VWFA; 1.
SQ SEQUENCE 902 AA; 99855 MW; F2BA57DAA5D4AAAD CRC64;

```

```

alignment_scores:
  Quality: 136.50      Length: 75
  Ratio: 2.625        Gaps: 2
Percent Similarity: 69.333 Percent Identity: 40.000

alignment_block:
US-09-049-696-13 x Q9EQR4
..
Align seg 1/1 to: Q9EQR4 from: 1 to: 902

1 GTGCGGCTCTGGAGAGTTAACGAGCAGCAGGAGAGTATACCCCA 50
|||||:||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||:
680 ValArgValGlnAlaArgLysAsnLysAlaArgLeuSer...LeuArgGl 695
51 GCAGAGTGAGGAGCACTGTACAVACCTGGCTGGATTCAGATGATGAAATAC 100
|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||:
695 nLysAsnLysSerLeuTyrIleProGlyTyrValGluAsnGlyLysIleV 712
101 AATGGAATCCACAGACCTGAAATTAATAAGGATGATGTTCAACACAAG 150
|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||:

```

```

712 alLeuAsnProProArgProAspValGlnGluGluAlaIleGluAlaThr 728
151 CAAGTGTGTTTTCAGCAAACTCTCGGAGGCTCATTTGTGGCTTCTGA 200
|||||:||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||:
729 ValGluAspPheAsnArgValThrSerGlyGlySerPheThrValSerGl 745
201 TGTCCCAAAATGTCCTCCATACCTGAT 225
|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||:
745 yAlaPro.....ProAsp 749
..
seq_name: sp_rodent:088860

```

seq_documentation_block:

```

ID Q88860 PRELIMINARY; PRT; 901 AA.
AC Q88860;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CHLORIDE CHANNEL CACC.
GN CLCA1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Romio L., Musante L., Cinti R., Moran O., Seri M., Galletta L.J.V.;
RT "Characterization of a murine gene homologous to the bovine CACC
  chloride channel."
RL Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AF052746; AAC35003.1;
DR MGS; MGI:1316732; Clcal.
DR InterPro: IPR002035; VWFA.
DR SMART: SM00327; VWA; 1.
DR PROSITE: PS50234; VWFA; 1.
SQ SEQUENCE 901 AA; 100038 MW; 13DF068BF3337AAB CRC64;

```

alignment_scores:

```

  Quality: 135.50      Length: 72
  Ratio: 2.606        Gaps: 2
Percent Similarity: 72.222 Percent Identity: 38.889

```

alignment_block:

```

US-09-049-696-13 x 088860
..
Align seg 1/1 to: 088860 from: 1 to: 901

19 GTTAACGCGCAGCAGCAGGAGAGTG.....ATACCCGACGAGTGG 59
|||||:||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||:
682 ValGlnAlaGlnArgAsnLysThrArgLeuSerLeuArgGlnLysAsnLy 698
60 AGCACTGTACATACCTGGCTGGATTCAGATGAGAAATACAAATGGAATC 109
|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||:
698 sSerLeuTyrIleProGlyTyrValGluAsnGlyLysIleValLeuAsnP 715
110 CACCAAGACCTGAAATTAATAGGATGATGTTCAACACAGCAAGCACTGTGT 159
|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||:
715 roProArgProAspValGlnGluGluAlaIleGluAlaThrValGluAsp 731
160 TTCAGCAGCAACATCCTCGGAGGCTCATTTGTGGCTTCTGATGTCCTCAA 209
|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||:
732 PheAsnArgValThrSerGlyGlySerPheThrValSerGlyAlaPro... 747
210 TGTCTCCCATACCTGAT 225
|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||: |||:|||||:
748 .....ProAsp 749
..
seq_name: sp_rodent:Q9QX15

```

seq_documentation_block:

```

ID Q9QX15 PRELIMINARY; PRT; 902 AA.
AC Q9QX15;

```

RX	MEDLINE=20012773; PubMed=10544033;
RT	Lee D., Ha S., Kho Y., Cho K., Baik M., Choi Y.,
RA	"Induction of mouse Ca(2+)-sensitive chloride channel 2 gene during
RT	involution of mammary gland."; 264:933-937(1999).
RL	Biochem. Biophys. Res. Commun. ;
DR	EMBL; AF108501; AAF12731.1; -.
DR	MGI; 1931471; C1ca2.
DR	InterPro: IPR002035; vWFA.
DR	SMART; SM00327; VWA; 1.
DR	PROSITE; PS50234; vWFA; 1.
SQ	SEQUENCE 902 AA; 99866 MW; 0FE746CFSFB3D07F CRC64;

```

alignment_scores:
  Quality: 130.50      Length: 75
  Ratio: 2.510         Gaps: 2
  Percent Similarity: 69.333      Percent Identity: 38.667

alignment_block:
  US-00-049-696-13 x Q9R070 ..
  Align seq 1/1 to: Q9R070 from: 1 to: 902

```

51 GCAGGTGGACCACTGTCATACCTGCCTGGATTGAGAATGATGAATAC 100
66 valalgalvalglnalaalaglysalasnilysaladlgaueuse...leuarigl 99
695 nlvasnlsaslerleutvrilleprogltyrvlgiluasngllvlarilev 712

101 AATGGAATCCACCAAGACCTGAAATTAAATAAGGATGATGTTCAACACAAG 150

12 11 10 9 8 7 6 5 4 3 2 1

151 CAAGTGTGTTTCAGCAGAACATCCTCGGGAGGGCCATTGTGGCTTCGGA 200

/29 ValGluAspPheAsnArgValThrSerGlyGlySerLeuThrValSerGI 743

201. TGTCCTCAATGCTCCCATACCTGAT 225

745 yAlaPro.....ProAsp 749

This Page Blank (uspto)

OM of: US-09-049-696-13 to: SwissProt_39:* out_format : pfs

Date: Mar 30, 2002 2:52 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=framer-n2p.model -US09049696/runat_28032002_145238_2085/app_query.fasta_1.12579
-o=/cgn2_1/USPOT_spo01/US09049696 -DEV=xlp
-DB=SwissProt_39 -QFMT=fastan -SUFFIX=resp -GAPOP=12.000
-CAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=200000000
-USER=US09049696 -CGN1_1_165 -NCEPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1
```

Search information block:

Query: US-09-049-696-13

Query length: 227

Database: SwissProt_39.*

Database sequences: 100059

Database length: 36664827

Search time (sec): 306.030000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
SwissProt_39:ECIC_BOVIN	147.00	295.72	9.4e-10	903	P54281 bos taurus (bovine).
SwissProt_39:PAXI_MOUSE	69.50	135.29	2.03	361	P09084 mus musculus (mouse).
SwissProt_39:YHIQ_YEAST	67.00	124.26	4.37	688	P38708 saccharomyces cerevisiae.
SwissProt_39:UBP7_YEAST	66.50	119.34	5.28	1071	P40453 saccharomyces cerevisiae.
SwissProt_39:HSF2_ARATH	66.00	125.43	5.53	468	P09545 arabidopsis thaliana.
SwissProt_39:Y4RM_RHISN	65.00	125.78	7.07	350	P55646 rhizobium sp. (strain).
SwissProt_39:ITB3_MOUSE	65.00	118.75	7.75	787	P05489 mus musculus (mouse).
SwissProt_39:YES_CANFA	64.50	120.95	8.53	539	P02823 canis familiaris (dog).
SwissProt_39:CND3_SCHPO	63.50	114.57	11.90	875	P10429 schizosaccharomyces pombe.
SwissProt_39:VA32_VACCV	63.00	123.68	12.00	270	P33849 variola virus (protein).
SwissProt_39:EP31_CHICK	63.00	122.77	12.14	300	P21055 vaccinia virus (strain).
SwissProt_39:YH07_MOUSE	63.00	112.47	13.87	984	P07494 gallus gallus (chicken).
SwissProt_39:PSB1_SCHPO	62.50	114.95	15.22	652	P05026 schizosaccharomyces pombe.
SwissProt_39:HRX_MOUSE	62.00	98.43	21.36	3866	P55200 mus musculus (mouse).
SwissProt_39:PSBB_CVACA	61.00	113.84	22.49	509	P01928 cyanidium caldarium.
SwissProt_39:YN04_YEAST	60.50	116.00	24.79	350	P53844 saccharomyces cerevisiae.
SwissProt_39:FLIH_CAEEL	60.50	104.91	28.60	1257	P34268 caenorhabditis elegans.
SwissProt_39:PRU_PIG	60.00	118.59	27.17	229	P01238 sus scrofa (pig).
SwissProt_39:VG27_BPMLS	60.00	115.27	28.36	336	P05234 mycobacteriophage 15.
SwissProt_39:ITB3_HUMAN	60.00	107.88	31.20	788	P05106 homo sapiens (human).
SwissProt_39:RBM6_HUMAN	60.00	104.81	32.47	1123	P78332 homo sapiens (human).
SwissProt_39:YCCF_ECOLI	59.50	121.29	29.74	148	P37065 escherichia coli.
SwissProt_39:PAXI_HUMAN	59.50	113.56	32.86	361	P15863 homo sapiens (human).
SwissProt_39:IDH_STRMO	59.50	112.82	33.18	393	P05940 streptococcus mutans.
SwissProt_39:B2AR_FELCA	59.50	112.29	33.41	418	P09455 felis silvestris catu.
SwissProt_39:GGAL_HUMAN	59.50	108.61	35.04	639	P09495 homo sapiens (human).
SwissProt_39:CBL_MOUSE	59.50	105.68	36.39	896	P22682 mus musculus (mouse).
SwissProt_39:RIP3_MOUSE	59.50	114.52	36.94	1024	P47434 mus musculus (mouse).
SwissProt_39:RUI7_ARATH	59.00	111.02	38.49	427	P42407 arabidopsis thaliana.
SwissProt_39:PSBB_MESVI	59.00	109.51	39.25	508	P09404 mesostigma viride. ph.
SwissProt_39:YH07_HUMAN	59.00	108.93	39.55	543	P07947 homo sapiens (human).
SwissProt_39:EPBA_XENLA	59.00	103.77	42.27	985	P01571 xenopus laevis (afric).
SwissProt_39:CYA7_MOUSE	59.00	102.82	42.80	1099	P51829 mus musculus (mouse).
SwissProt_39:SHBG_PHOSU	59.00	110.52	43.91	399	P06258 photodupus sungorus (st).
SwissProt_39:TCMO_HELTU	59.00	108.48	45.09	505	P04468 helianthus tuberosus.
SwissProt_39:ICP4_HSV11	58.50	108.48	45.09	505	P04320 zinnia elegans. trans.
SwissProt_39:COX3_PELSU	58.50	100.29	50.12	1298	P08392 herpes simplex virus.
SwissProt_39:NIF4_CLOPA	58.00	113.11	48.13	261	P07967 pelomeda subrufa (a).
SwissProt_39:Y4NJ_RHISN	58.00	112.72	48.38	273	P22548 clostridium pasteurian.
SwissProt_39:Y4NJ_RHISN	58.00	107.39	51.83	505	P55582 rhizobium sp. (strain).

alignment_scores:

Quality: 147.00

Length: 63

SwissProt_39:TNSD_ECOLI + 58.00 107.34 51.86 508 | P13991 escherichia coli.
SwissProt_39:Y262_HAEIN + 58.00 104.28 53.96 723 | P44600 haemophilus infue
SwissProt_39:HXK2_HAEIN + 58.00 104.25 53.97 725 | P45357 haemophilus infue
SwissProt_39:EPBB_XENLA + 58.00 102.36 55.31 902 | Q91736 xenopus laevis (af
SwissProt_39:YJ11_YEAST - 58.00 100.07 56.97 1174 | P47108 saccharomyces cer

seq_name: SwissProt_39:ECIC_BOVIN

seq_documentation_block:

ID ECIC_BOVIN STANDARD; PRT; 903 AA.
AC P54281;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EPITHELIAL CHLORIDE CHANNEL PROTEIN (CALCIUM-ACTIVATED CHLORIDE
DE CHANNEL).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Trachea;
RX MEDLINE=96125078; PubMed=8537359;
RA Cunningham S.A., Awayda M.S., Bublén J.K., Ismailov I.I.,
RA Arrate M.P., Berdiev B.K., Benos D.J., Fuller C.M.;
RT "Cloning of an epithelial chloride channel from bovine trachea.";
RL J. Biol. Chem. 270:31016-31026(1995).
CC -!- FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE
CC SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME;
CC MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND
CC TRANSEPITHELIAL ION TRANSPORT.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- TISSUE SPECIFICITY: TRACHEA.
CC -!- PTM: PHOSPHORYLATED BY CAM-KINASE II.
CC -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U36445; AAC48511.1; -
InterPro: IPR002035; VWFA.
SMART: SM00327; VWFA; 1.
PROSITE: PS0234; VWFA; 1.

Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
Calcium channel; Phosphorylation;
TRANSMEM 7 27 POTENTIAL.

TRANSMEM 331 351 POTENTIAL.

TRANSMEM 617 637 POTENTIAL.

TRANSMEM 883 903 POTENTIAL.

DOMAIN 308 476 VWFA.

CARBOHYD 75 75 N-LINKED (GLCNAC. .) (POTENTIAL).

CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).

CARBOHYD 360 360 N-LINKED (GLCNAC. .) (POTENTIAL).

CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).

CARBOHYD 504 504 N-LINKED (GLCNAC. .) (POTENTIAL).

CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).

CARBOHYD 688 688 N-LINKED (GLCNAC. .) (POTENTIAL).

CARBOHYD 811 811 N-LINKED (GLCNAC. .) (POTENTIAL).

CARBOHYD 816 816 N-LINKED (GLCNAC. .) (POTENTIAL).

CARBOHYD 842 842 N-LINKED (GLCNAC. .) (POTENTIAL).

CARBOHYD 857 857 N-LINKED (GLCNAC. .) (POTENTIAL).

SEQUENCE 903 AA; 100305 MW; 9742AB1590908AEC CRC64;

Ratio: 3.196 Gaps: 2
Percent Similarity: 73.016 Percent Identity: 52.381

alignment_block:

US-09-049-696-13 x ECLC_BOVIN ..

Align seg 1/1 to: ECLC_BOVIN from: 1 to: 903

22 AACGACGACGAGGAGTATACCCAGCAGAGTGAGGACCTGATCAT 71
||||:||||| :|||: |||||
689 AsnThAlaArgLeuSerLeuArgGlnProGlnAsnLysAlaLeuTyr 705
72 ACCTGCTGATGAGATGATGAAATCAATGGAATCCACCAAGACCTG 121
||||:||||| :|||: |||||
705 eProGlyTyrIleGluAsnGlyLysIleIleLeuAsnProArgProG 722
122 AAATTAATAGGATGATGTCACACCAAGAGTG...TGTTTCAGCAGA 168
||||:||||| :|||: |||||
722 luVal...LysAspAspLeuAlaLysAlaGluIleGluAspPheSerArg 737
169 ACATCTCGGAGGCTCATTTGGCTTCTGATGTCCTCA 207
:::|||||:|||||:|||||:|||||
738 LeuthSerGlyGlySerPheThrValSerGlyAlaPro 750

seq_name: SwissProt_39:PAX1_MOUSE

seq_documentation_block:

ID PAX1_MOUSE STANDARD; PRT; 361 AA.
AC P09084; Q9R2B1;
DT 01-NOV-1988 (Rel. 09, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PAIRED BOX PROTEIN PAX-1.
GN PAX1 OR PAX-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91364170; PubMed=1889089;
RA Chalepakis G., Fritsch R., Fickenscher H., Deutsch U., Goulding M.D.,
RA Gruss P.;
RT "The molecular basis of the undulated/Pax-1 mutation.";
RL Cell 66:873-884(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RA Egger C., Schorpp M., Boehm T.;
RT "Evolutionary conservation of gene structure of the pax1/9 gene
family.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 6-132 FROM N.A.
RX MEDLINE-88223354; PubMed=2453291;
RA Deutsch U., Dressler G.R., Gruss P.;
RT "Pax 1, a member of a paired box homologous murine gene family, is
expressed in segmented structures during development.";
RL Cell 53:617-625(1988).
RN [4]
RP VARIANT UNDATED.
RX MEDLINE-89028662; PubMed=3180219;
RA Balling R., Dressler G.R., Gruss P.;
RT "Undulated, a mutation affecting the development of the mouse
skeleton, has a point mutation in the paired box of Pax 1.";
RL Cell 55:531-535(1988).
CC -!- FUNCTION: THIS PROTEIN IS A TRANSCRIPTIONAL ACTIVATOR. IT MAY PLAY
A ROLE IN THE FORMATION OF SEGMENTED STRUCTURES OF THE EMBRYO. MAY
PLAY AN IMPORTANT ROLE IN THE NORMAL DEVELOPMENT OF THE VERTEBRAL
COLUMN.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: UNDATED (UN) HOMOEZYGOUS MICE EXHIBITS VERTEBRAL
MALFORMATIONS ALONG THE ENTIRE ROSTRO-CAUDAL AXIS. THIS IS

CC DUE TO A SINGLE MUTATION IN THE PAIRED BOX REGION.
CC -!- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M69222; AAA39888.1; ..
DR EMBL; AJ133503; CAB38370.1; ..
DR EMBL; AJ133504; CAB38370.1; JOINED.
DR EMBL; AJ133505; CAB38370.1; JOINED.
DR EMBL; M20978; AAA37793.1; ALT_SEQ.
DR PIR; A29902; A29902.
DR PIR; A40023; A40023.
DR HSP; P06601; IPDN.
DR TRANSFAC; T00677; ..
DR MGD; MGI:97485; Pax1.
DR InterPro; IPR001523; Paired_box.
DR Pfam; PF00292; PAX; 1.
DR PRINTS; PR00027; PAIREDBOX.
DR SMART; SM00351; PAX; 1.
DR PROSITE; PS00034; PAIRED_BOX; 1.
KW Paired box; Developmental protein; Nuclear protein; DNA-binding;
KW Transcription regulation; Activator; Disease mutation.
FT DCMAIN 4 128 PAIRED BOX.
FT VARIANT 18 18 G -> S (IN UNDATED).
FT CONFLICT 56 56 I -> M (IN REF. 3).
FT CONFLICT 215 215 A -> T (IN REF. 1).
FT CONFLICT 260 260 S -> L (IN REF. 1).
FT CONFLICT 282 282 H -> Q (IN REF. 1).
SQ SEQUENCE 361 AA; 37859 MW; FCB69BE6D51544AB CRC64;

alignment_scores:
Quality: 69.50 Length: 76
Ratio: 1.580 Gaps: 5
Percent Similarity: 57.895 Percent Identity: 38.158

alignment_block:
US-09-049-696-13/rev x PAX1_MOUSE ..
Align seg 1/1 to: PAX1_MOUSE from: 1 to: 361

212 GCATTGGGACATCAGAACCCACAAAT.....GAGCCTCCGAGGA 172
||||| ||||| ||||| |||||
235 AlaPheProThrSerProAlaValAsnGlyLeuGluLysProAlaLeuG 252

171 TGTCTGCTGAACACACACTGCTGTGTGTGAACATCATCTTATTATT 122
::: |||||:|||||
252 uAlaAspIleLysThr.....GlnSerAlaSerSerLeu 265

121 CAGGTCTTGGTGGATTC.....CATTTGATTTCATCATCTCAATCCAG 78
||:|||||:||||| |||||
265 eAlaValGlyGlyPheLeuProAlaCysAlaTyrProAlaSerAsnGln 281

77 CCAGGTATGATACAGTCTCCACTCTGCTGGGTATCATCTCCGTCTGCG 28
||:|||||:||||| |||||
282 HisGlyValTyrSerAlaProAlaAla.GlyTyrLeuSerProGlyPro 297

27 TGC GTTAACTCTCCAGAGCCCG 4
298 ProTyrProProAlaGlnAlaPro 305

seq_name: SwissProt_39:YH10_YEAST
seq_documentation_block:
ID YH10_YEAST STANDARD; PRT; 688 AA.
AC P38708;
DT 01-FEB-1995 (Rel. 31, Created)


```

569 AsnGlnValValValArgAsnAspSerLysLysTyrValValVal 565
193 T...GATGTCCTCCCAATGCTCCCATCTGAT 225
565 rPheaspGluLeuGluAlaArgIleProGlu 575

seq_name: SwissProt_39:UBP7_YEAST
seq_documentation_block:
ID AD UBEP7_YEAST STANDARD; PRT; 1071 AA.
AC F40453;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 7 (EC 3.1.2.15) (UBIQUITIN
DE THIOLESTERASE 7) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 7)
DE DE (UBIQUITININATING ENZYME 7).
DE DE UBEP7 OR YIL156W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowden S., Brown D.,
RA Church C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagers K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skellton J., Smith V.,
RA Walsh S.V., Whitehead S.;
PL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O -
CC UBIQUITIN + A THIOL.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19; ALSO KNOWN AS
CC -1- FAMILY 2 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
EMBL: Z38059; CAA86122.1; -.
DR PIR: S48378; S48378.
DR SGD: S0001418; UBEP7.
DR InterPro: IPR001763; Rhodanese_domain.
DR InterPro: IPR001394; UCH-2.
DR Pfam: PF00581; Rhodanese; 1.
DR Pfam: PF00442; UCH-1; 1.
DR Pfam: PF00443; UCH-2; 1.
DR SMART: SM00450; RHOD; 1.
DR PROSITE: PS00972; UCH_2_1; 1.
DR PROSITE: PS00973; UCH_2_2; 1.
DR PROSITE: PS50235; UCH_2_3; 1.
DR Ubiquitin conjugation; Hydrolase; Thiol protease; Multigene family.
KW ACT_SITE 618 618
FT ACT_SITE 1006 1006 BY SIMILARITY.
FT ACT_SITE 1014 1014 BY SIMILARITY.
FT ACT_SITE 1014 1014 BY SIMILARITY.
SQ SEQUENCE 1071 AA; 123133 MW; 82683A01063CEC74 CRC64;

alignment_scores:
Quality: 66.50 Length: 54
Ratio: 2.145 Gaps: 2
Percent Similarity: 57.407 Percent Identity: 31.481

alignment_block:
US-09-049-696-13/rev x UBEP7_YEAST ..

```

Tue Apr 2 09:39:42 2002

Align seg 1/1 to: UBP7_YEAST from: 1 to: 1071

```
226 GATCAGTATGGAGCATTGGACATCAGAACGACCAAAATGAGCTCCC 177
|||||
31 AspLeuTyHisGluAspIleHisTyTyProGlnLeuLysLeuG1 47
|||||
176 GAGGATGTTCTGCTGAAACACACTTGTGTTTGAACATCATCTTATT 127
|||||
47 uLysLeu.....LeuAspLeuLeuGluHisThrGluTyrL 59
|||||
126 AATTCAGCTTGTGGGATTCATGTTATTCATCATCTCAATCCAGC 77
|||||
59 euPheGluLeuTyTyLeuAspSerIle.....HisHisaspargProAsn 73
|||||
76 CAGGTATGTACA 65
|||||
74 AspAlaLeuThr 77
```

seq_name: SwissProt_39:HSF2_ARATH

```
seq_documentation_block:
ID HSF2_ARATH STANDARD; PRT; 468 AA.
AC Q9SCW5;
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HEAT SHOCK FACTOR PROTEIN 2 (HSF 2) (HEAT SHOCK TRANSCRIPTION FACTOR
DE 2) (HSTF 2).
OS Arabidopsis thaliana (Mouse-ear cress).
GN Arabidopsis thaliana (Mouse-ear cress);
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OC NCBI_TaxID=3702;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=CV. COLOMBIA; TISSUE=Green siliques;
RA Schoeffli F., Praendel R.;
RT "De-repression of heat shock protein synthesis in transgenic plants.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBSJ databases.
CC -!- FUNCTION: DNA-BINDING PROTEIN THAT SPECIFICALLY BINDS HEAT SHOCK
CC PROMOTER ELEMENTS (HSE) AND ACTIVATES TRANSCRIPTION (BY
CC SIMILARITY).
CC -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- PTM: EXHIBITS TEMPERATURE-DEPENDENT PHOSPHORYLATION
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HSF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ251865; CAB63800.1; .
CC HSSP; P22121; 3HSF.
CC InterPro; IPR000232; HSF_DNA_bind.
CC InterPro; IPR002341; HSF_ETS.
CC Pfam; PF00447; HSF_DNA-bind; 1.
CC PRINTS; PR000556; HSFDOMAIN.
CC ProDom; PD001788; HSF_DNA_bind; 1.
CC SMART; SM00415; HSF; 1.
CC Transcription regulation; Nuclear protein; DNA-binding; Activator;
KW phosphorylation; Heat shock; Multigene family.
FT DNA_BIND 21 115 BY SIMILARITY.
SQ SEQUENCE 468 AA; 51949 MW; 4F1184C8AB8B1F92 CRC64;
```

alignment_scores: Quality: 66.00 Length: 79

Ratio: 1.833 Gaps: 4
Percent Similarity: 45.570 Percent Identity: 30.380

alignment_block:

US-09-049-696-13 x HSF2_ARATH ..

Align seg 1/1 to: HSF2_ARATH from: 1 to: 468

```
16 GGAGTTAAACGACGACGAGAGTATACCCAGCAGAGTGGAGCACT 65
|||||
240 GlyValAsnGlyLeuSerArgGlnIleValArgTyGlnSerSerMet.. 255
|||||
66 GTACATACCTGGCTGGATTCGAGAATGAT.....G 94
|||||
256 .....AsnAspAlaThrAsnThrMetLeuGlnG 265
|||||
95 AATACACAAATGG.....AATCCACCAAGACCTGAAATTAATAGGATGAT 138
|||||
265 InileGlnGlnMetSerAsnAlaProSerHisGlu..... 276
|||||
139 GTTCAACACAAAGCAAGTGTGTTTCAGCAGAACATCTCTCGGAGGCTCATT 188
|||||
277 .....SerLeuSerSerAsnAsnGlySerPh 285
```

seq_name: SwissProt_39:Y4RM_RHISN

seq_documentation_block:

```
ID Y4RM_RHISN STANDARD; PRT; 350 AA.
AC P5646;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 39.4 KDA PROTEIN Y4RM.
GN Y4RM.
OS Rhizobium sp. (strain NGR234).
OC plasmid sym. NGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OC NCBI_TaxID=394;
OX [1]
RN [1]
RS SEQUENCE FROM N.A.
RC MEDLINE=97305956; PubMed=9163424;
RA Friberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -!- SIMILARITY: NONE OBVIOUS OUTSIDE OF SIMILARITY WITH Y4ZA.
CC -!- POTENTIAL FRAGMENT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000094; AAB91838.1; .
CC InterPro; IPR001584; Rve.
CC Pfam; PF00665; rve; 1.
CC Hypothetical protein; Plasmid.
FT SIMILAR 258 339 HIGHLY SIMILAR TO Y4ZA 34-115.
SQ SEQUENCE 350 AA; 39381 MW; 6CBE6FACADD70DFC CRC64;
```

alignment_scores: Quality: 65.00 Length: 68
Ratio: 1.857 Gaps: 3
Percent Similarity: 51.471 Percent Identity: 32.353

alignment_block:

US-09-049-696-13 x Y4RM_RHISN

Align seg 1/1 to: Y4RM_RHISN from: 1 to: 350

4 CGGGCTGGGAGGAGTTAAGCGAGCGAGCGGAGAGTGATACCCAGCA 53
142 ArgArgArgAlaGlyMetSerSerAlaValArgSerValProValar 158
54 GAGTGGAGCACTGTACATACCTGGCTGGATTGAGATGATGAATACAAT 103
158 gThrPhegly.....AspT 163
104 GGAATCCACCAAGACCT...GAATTAATAGGATGATGTTCAACACAAG 150
163 rPasnAspProLeuProGlyTyrValGluValaspPheValalaHis... 178
151 CAAGTGTGTTTCAGCAGCAACATCTCGGAGGCTCATTTGGTCTTCGA 200
179SerGlyThrSerSerSerGlySerPheValGlnThrIle 191
201 TGTC 204
191 tVal 192

seq_name: SwissProt_39:ITB3_MOUSE

seq_documentation_block:

ID ITB3_MOUSE STANDARD; PRT: 787 AA.
AC O54890;
DT 15-DEC-1998 (Rel. 37, Created)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTEGRIN BETA-3 PRECURSOR (PLATELET MEMBRANE GLYCOPROTEIN IIIA)
DE (GPIIIA) (CD61).
GN ITGB3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA McHugh K.P.; Teitelbaum S.L.; Kitazawa S.; Ross F.P.;
RL Submitted (SEP-1987) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INTEGRIN ALPHA-V/BETA-3 IS A RECEPTOR FOR CYTOTOXICIN,
CC FIBRONECTIN, LAMININ, MATRIX METALLOPROTEINASE-2, OSTEOPOINTIN,
CC PROTHROMBIN, THROMBOSPONDIN, VITRONECTIN AND VON WILLEBRAND
CC FACTOR. INTEGRIN ALPHA-IIB/BETA-3 IS A RECEPTOR FOR FIBRONECTIN,
CC FIBRINOGEN, PLASMINOGEN, PROTHROMBIN, THROMBOSPONDIN AND
CC VITRONECTIN. INTEGRINS ALPHA-IIB/BETA-3 AND ALPHA-V/BETA-3
CC RECOGNIZE THE SEQUENCE R-G-D IN A WIDE ARRAY OF LIGANDS. INTEGRIN
CC ALPHA-IIB/BETA-3 RECOGNIZES THE SEQUENCE H-H-G-G-A-K-Q-A-G-D-V
CC IN FIBRINOGEN GAMMA CHAIN. FOLLOWING ACTIVATION INTEGRIN ALPHA-
CC IIB/BETA-3 BRINGS ABOUT PLATELET/PLATELET INTERACTION THROUGH
CC BINDING OF SOLUBLE FIBRINOGEN. THIS STEP LEADS TO RAPID PLATELET
CC AGGREGATION WHICH PHYSICALLY PLUGS RUPTURED ENDOTHELIAL SURFACE.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-3
CC ASSOCIATES WITH EITHER ALPHA-IIB OR ALPHA-V.
CC -!- PTM: PHOSPHORYLATED ON TYROSINE RESIDUES IN RESPONSE TO THROMBIN-
CC INDUCED PLATELET AGGREGATION. PROBABLY INVOLVED IN OUTSIDE-IN
CC SIGNALING (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.

CC EMBL; AF026509; AAB94086.1; -
CC MGD; MGI:96612; Itgb3.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002369; Integrin_B.
DR InterPro: IPR001189; Integrin_Beta_C.
DR InterPro: IPR003639; PSI.
DR InterPro: IPR002035; VWFA.
DR Pfam: PF00362; Integrin_B; 1.
DR PRINTS; PRO1186; INTEGRINB.
DR ProDom: PD001811; Integrin_B; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00187; INB; 1.
DR SMART; SM00423; PSI; 1.
DR SMART; SM00327; VWFA; 1.
DR PROSITE; PS00243; INTEGRIN_BETA; 3.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
KW Repeat; Extracellular matrix; Cytoskeleton; Signal; Phosphorylation.
ET SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 787 INTEGRIN BETA-3.
FT DOMAIN 26 717 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 718 740 POTENTIAL.
FT DOMAIN 741 787 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 134 376 VWFA-LIKE.
FT DOMAIN 462 628 4 CYSTEINE-RICH TANDEM REPEATS.
FT REPEAT 511 552 I.
FT REPEAT 553 591 II.
FT REPEAT 592 628 III.
FT DISULFID 30 460 IV. BY SIMILARITY.
FT DISULFID 38 48 BY SIMILARITY.
FT DISULFID 41 74 BY SIMILARITY.
FT DISULFID 51 63 BY SIMILARITY.
FT DISULFID 202 209 BY SIMILARITY.
FT DISULFID 257 298 BY SIMILARITY.
FT DISULFID 399 411 BY SIMILARITY.
FT DISULFID 431 680 BY SIMILARITY.
FT DISULFID 458 462 BY SIMILARITY.
FT DISULFID 473 485 BY SIMILARITY.
FT DISULFID 482 520 BY SIMILARITY.
FT DISULFID 487 496 BY SIMILARITY.
FT DISULFID 498 511 BY SIMILARITY.
FT DISULFID 526 531 BY SIMILARITY.
FT DISULFID 528 561 BY SIMILARITY.
FT DISULFID 533 546 BY SIMILARITY.
FT DISULFID 548 553 BY SIMILARITY.
FT DISULFID 567 572 BY SIMILARITY.
FT DISULFID 569 600 BY SIMILARITY.
FT DISULFID 574 583 BY SIMILARITY.
FT DISULFID 585 592 BY SIMILARITY.
FT DISULFID 606 611 BY SIMILARITY.
FT DISULFID 608 656 BY SIMILARITY.
FT DISULFID 613 623 BY SIMILARITY.
FT DISULFID 626 629 BY SIMILARITY.
FT DISULFID 633 642 BY SIMILARITY.
FT DISULFID 639 712 BY SIMILARITY.
FT DISULFID 660 688 BY SIMILARITY.
FT CARBOHYD 345 345 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 396 396 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 584 584 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 679 679 N-LINKED (GLCNAC...) (POTENTIAL).
FT MOD_RES 772 772 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 784 784 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 787 AA; 86694 MW; B1570599ABC438A3 CRC64;

alignment_scores:

Quality: 65.00 Length: 40
Ratio: 2.708 Gaps: 2


```
DR InterPro: IPR001090; Ephrin_rcptor.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001777; FN_III.
DR InterPro: IPR001426; Raptor_tyr_kin_v.
DR InterPro: IPR001660; SAM.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam: PF01404; EPH_lbd; 1.
DR Pfam: PF00041; fn3; 4.
DR Pfam: PF00089; pkinase; 2.
DR Pfam: PF00536; SAM; 2.
DR ProDom: PD001495; Ephrin_rcptor; 1.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00454; SAM; 1.
DR SMART: SM00219; Tykrc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00790; RECEPTOR_TYR_KIN_V_1; PARTIAL.
DR PROSITE: PS00791; RECEPTOR_TYR_KIN_V_2; 1.
DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Receptor; Transmembrane; Glycoprotein; Repeat.
FT NON_TER 1 1 EXTRACELLULAR (POTENTIAL).
FT DOMAIN <1 541 POTENTIAL.
FT TRANSMEM 542 562 POTENTIAL.
FT DOMAIN 563 984 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 164 300 CYS-RICH.
FT DOMAIN 301 410 FIBRONECTIN TYPE-III 1.
FT DOMAIN 411 527 FIBRONECTIN TYPE-III 2.
FT DOMAIN 619 882 PROTEIN KINASE.
FT DOMAIN 909 984 SAM.
FT SITE 982 984 PDZ-BINDING MOTIF (POTENTIAL).
FT NP_BIND 625 633 ATP (BY SIMILARITY).
FT BINDING 651 651 ATP (BY SIMILARITY).
FT ACT_SITE 744 744 BY SIMILARITY.
FT MOD_RES 594 594 PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT MOD_RES 600 600 PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT MOD_RES 778 778 PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT MOD_RES 928 928 PHOSPHORYLATION (AUTO-) (POTENTIAL).
SQ SEQUENCE 984 AA; 109519 MW; EF06C83B863A13A1 CRC64;

alignment_scores:
  Quality: 63.00 Length: 21
  Ratio: 3.938 Gaps: 0
Percent Similarity: 76.190 Percent Identity: 47.619

alignment_block:
US-09-049-696-13 x EPBL_CHICK ..
Align seg 1/1 to: EPBL_CHICK from: 1 to: 984

97 ATACATGGAATCCACCAAGACCTGAATTAAGATGATGTTCAACA 146
   :|||||:|||||:
321 LeuGluIrpAsnProProArgGluThrGlycylArgAspValThrTy 337
   :|||||:

147 CAACGAAGTGTTG 159
   :|||:
337 rAsnIleValCys 341

seq_name: SwissProt_39:SSP1_SCHPO

seq_documentation_block:
ID SSP1_SCHPO STANDARD; PRT; 652 AA.
AC P50526;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SERINE/THREONINE-PROTEIN KINASE SSP1 (EC 2.7.1.-).
GN SSP1 OR SPCC297.03.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
```

```
CC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
PC STRAIN=972;
RX MEDLINE=95354651; PubMed=7628434;
RA Matsusaka T., Hirata D., Yanagida M., Toda T.;
RT "A novel protein kinase gene sspl+ is required for alteration of
RL growth polarity and actin localization in fission yeast.";
RL EMBO J. 14:3325-3338(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Furnelle B., Goffeau A., Lyne M., Rajandream M.A., Barrell B.G.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -1: FUNCTION: INVOLVED IN ACTIN LOCALIZATION AND THUS IN POLARIZED
CC CELL GROWTH.
CC -1: SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -1: SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -1:
DR EMBL: D45882; BAA08301.1;
DR EMBL: AL049609; CAB40783.1;
DR HSSP: Q63450; IA06.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR Pfam: PF00069; pkinase; 1.
DR SMART: SM00220; S_TKc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 135 409 PROTEIN KINASE.
FT NP_BIND 141 149 ATP (BY SIMILARITY).
FT BINDING 164 164 ATP (BY SIMILARITY).
FT ACT_SITE 267 267 BY SIMILARITY.
SQ SEQUENCE 652 AA; 73992 MW; B1931E5EB75A85DA CRC64;

alignment_scores:
  Quality: 62.50 Length: 28
  Ratio: 2.717 Gaps: 1
Percent Similarity: 82.143 Percent Identity: 42.857

alignment_block:
US-09-049-696-13 x SSP1_SCHPO ..
Align seg 1/1 to: SSP1_SCHPO from: 1 to: 652

89 GGATTGGAATGATGAATACAAATCGA...ATCCACCAAGACCTGAAT 126
   :|||||:
254 GlyLeuGluTyrLeuHisTyrGlnGlyIleIleHisArgAspIleLysPr 270
   :|||||:

127 AATAAGGATGATGTTCAACACACAAAGTGTTGTT 160
   :|||||:
270 oAlaAsnLeuLeuAsnSerSerAsnCysVal 281

seq_name: SwissProt_39:HRX_MOUSE

seq_documentation_block:
ID HRX_MOUSE STANDARD; PRT; 3866 AA.
AC P55200;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ZINC FINGER PROTEIN HRX (ALL-1) (FRAGMENT).
OC
```


CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF022186; AAB82661.1; -
DR InterPro; IPR000932; PSII.
DR Pfam; PF0421; PSII; 1.
KW Photosynthesis; Photosystem II; Thylakoid; Chlorophyll; Chloroplast;
KW Transmembrane
FT TRANSMEM 19 40 POTENTIAL.
FT TRANSMEM 95 118 POTENTIAL.
FT TRANSMEM 138 163 POTENTIAL.
FT TRANSMEM 197 219 POTENTIAL.
FT TRANSMEM 236 260 POTENTIAL.
FT TRANSMEM 449 468 POTENTIAL.
SQ SEQUENCE 509 AA; 56561 MW; D786C038AD3217D6 CRC64;

alignment_scores:
Quality: 61.00 Length: 61
Ratio: 1.525 Gaps: 4
Percent Similarity: 65.574 Percent Identity: 34.426

alignment_block:

US-09-049-696-13/rev x PSBB_CYACA ..

Align seg 1/1 to: PSBB_CYACA from: 1 to: 509

196 AAGCCACAATGAGCCTCCGAGGATGTTCTGCTGAACACACTTGCTTG 147
::||| ::|||::|||::|||::|||
130 GluProAlaLeuAspLeuProLysValPheGly..IleHisLeuValLeu 145
146 TGTGACATCATCTTATTATTAATTCAGGTCTTGCTGGATTCATTCTAT 97
|||::|||::|||::|||::|||::|||
146SerSerLeuLeuCysPheGlyAlaPheHisValTh 159
96 TTCATCATCTCAATCCAGCAGGATGATACAGTGCCTCCACTCTGCTGG 47
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
159 rGlyLeuPheGly.....ProGlyIleTrpIleSerAsp...AlaTyrg 173
46 GTATCACTCTCCGCTGCTGGCTGCGTTAACTCCT 15
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
173 lyLeuThrGlyArgIleGlnSerValAlaPro 183

This Page Blank (uspto)

OM of: US-09-049-696-13 to: PIR_58:* out_format : pfs

Date: Mar 30, 2002 2:27 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09049696/runat_28032002_145238_2031/app_query.fasta_1.12579
-DB=PIR_68 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09049696_@CGN1_1_367 -NCPU=6 -ICPU=3
-LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-049-696-13

Query length: 227

Database: PIR_68:

Database sequences: 219241

Database length: 76174552

Search time (sec): 470.790000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
pir2:G0168	+	215.50	458.36	1.7e-18	913	! gob-5 protein - mouse
pir2:T02205	+	144.00	297.54	1.5e-09	905	! Lu-ECAM-1 protein - bovine
pir2:A40023	+	71.50	142.22	1.73	361	! paired box homolog PAX1 - mouse
pir2:S46774	+	67.00	126.60	6.73	688	! multifunctional amino acid--trn
pir2:S48378	-	66.50	121.70	8.10	1071	! probable membrane protein Y11
pir2:J07319	-	66.00	129.94	8.45	357	! probable allatostatin receptor
pir2:J07862	+	66.00	124.94	8.92	642	! receptor-like protein kinase PR
pir2:J48004	+	65.00	124.32	11.70	530	! multifunctional aminoacyl--trna
pir2:PN0510	+	65.00	122.20	11.97	680	! integrin beta-3 chain - mouse
pir2:PN0509	+	65.00	121.67	12.04	723	! integrin beta-3 chain - rat (fr
pir2:PN04812	-	64.50	127.34	12.92	326	! hypothetical protein A2g39050
pir2:A70906	+	64.50	126.26	13.07	370	! probable RNA polymerase sigma s
pir2:T25190	+	63.50	127.59	16.78	243	! hypothetical protein r23G11.5
pir2:T43522	+	63.50	116.67	18.90	875	! condensin complex component and
pir2:T37418	+	63.00	125.60	19.56	269	! probable 30.8K protein - vaccin
pir2:T28577	+	63.00	125.57	19.57	270	! ATP/GTP-binding protein A35L -
pir2:T36851	+	63.00	125.57	19.57	270	! SalL2L protein - vaccinia virus
pir2:J01768	+	63.00	125.57	19.57	270	! A36L protein - variola minor vi
pir2:A72168	+	63.00	124.67	19.76	300	! A32L protein - vaccinia virus
pir2:T42520	+	63.00	122.09	20.33	406	! probable oxidoreductase - Strept
pir2:T36632	+	63.00	119.12	21.00	575	! hypothetical protein BH2474 (im
pir2:B83959	+	63.00	117.51	21.37	695	! hypothetical protein YL0035c -
pir2:S64862	+	63.00	114.83	22.00	952	! serine/threonine kinase (EC 2.7
pir2:T50612	+	62.50	116.93	24.54	652	! major facilitator family protein
pir2:S58666	+	62.00	118.81	27.43	458	! cytosine methylase-- phage Hp1
pir2:B75413	+	62.00	114.76	28.67	737	! hypothetical protein G01D9.2 -
pir2:S38866	+	62.00	100.63	33.45	3869	! All-1 protein +GTE form - mous
pir2:T31349	+	62.00	100.63	33.45	3869	! All-1 protein +GTE form - mous
pir2:A48205	+	61.50	123.94	29.60	220	! hypothetical protein F14D17.10
pir2:T48970	+	61.50	110.58	34.25	1055	! multidrug-efflux transporter X
pir2:C82600	-	61.50	101.58	37.78	3034	! seven-pass transmembrane rece
pir2:T14119	-	61.00	129.20	31.90	104	! hypothetical protein DXF2761E1
pir2:T46246	-	61.00	120.37	35.12	293	! hypothetical protein PA5270 (im
pir2:B82988	-	61.00	115.66	36.97	509	! photosystem II p680 chlorophyl
pir2:T11996	-	60.50	121.05	39.78	357	! hypothetical protein T09F5.6 -
pir2:T24759	-	60.50	117.73	41.25	230	! probable membrane protein YN26
pir2:S60917	+	60.50	116.09	41.99	424	! hypothetical protein 1 - Caenor
pir2:S27783	+	60.50	112.94	43.47	614	! probable protein kinase (Import
pir2:B86461	+	60.50	110.18	44.79	849	! hypothetical protein F5A12.5 -
pir2:T15190	+	60.00	128.77	41.74	84	! hypothetical protein - Synchoch

pir2:A82466 + 60.00 121.33 45.26 201 ! hypothetical protein VCA0380
pir1:ICPG - 60.00 120.22 45.81 229 ! prolactin precursor - pig
pir2:D84001 + 60.00 118.63 46.62 276 ! lipote protein ligase BH281
pir2:S30972 + 60.00 116.95 47.48 336 ! minor tail protein - Mycobac
seq_name: pir2:JG0168
seq_documentation_block:
gob-5 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 11-May-2000
C:Accession: JG0168
R:Komiyama, T.; Tanigawa, Y.; Hirohashi, S.
Biochem. Biophys. Res. Commun. 255, 347-351, 1999
A:Title: Cloning and identification of the gene gob-5, which is expressed in intestine
A:Reference number: JG0168; MUID:99160866
A:Accession: JG0168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-913 <KOM>
A:Cross-references: DDBJ:AB016592
alignment_scores:
Quality: 215.50 Length: 74
Ratio: 3.716 Gaps: 2
Percent Similarity: 78.378 Percent Identity: 59.459
alignment_block:
US-09-049-696-13 x JG0168 ..
Align seg 1/1 to: JG0168 from: 1 to: 913
7 GCTCGGAGGAGTAAACGACGACGAGAGTATACCCAGCAGAG 56
|||||
673 AlaLeuGlyGlyValThrSerAspArgGlnArgAlaAProProLysAs 694
57 TGGAGCACTGTACATACCTGCTGGTGGTGAATGACGAATACATGGA 106
: |||||
694 nArgAlaMetTyrIleAspGlyTyrPileGluAspGlyGluValArgMeta 711
|||||
107 ATCCACCAAGACCTGAATTAATAGGATGATTTCAACACCAAGCAAGTG 156
|||||
711 snProProArgProGluThrSerTyr.....ValGlnAspLysGlnLeu 725
157 TGTTCAGCAGAACATCTCGGAGGCTCATTTGTGGCTTCTGATGTCCC 206
|||||
726 CysPheSerThrArgSerSerGlySerPheValAlaThrAsnValPr 742
207 A....AATGCTCCCATCTGAT 225
|||||
742 oAlaAlaAProIleProAsp 749

seq_name: pir2:T02205
seq_documentation_block:
Lu-ECAM-1 protein - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 01-Dec-2000
C:Accession: T02205; T02152; T02171
R:Elble, R.C.; Widom, J.; Gruber, A.D.; Abdel-Chany, M.; Levine, R.; Goodwin, A.; Pau
submitted to the EMBL Data Library, April 1997
A:Description: Cloning and characterization of Lu-ECAM-1 suggest it is an endothelial
A:Reference number: Z14590
A:Accession: T02205
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-905 <ELB>
A:Cross-references: EMBL:AF001261; NID:g2623762; PIDN:AB86529.1; PID:g2623763
A:Experimental source: lung
A:Accession: T02152
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA


```

49 .....CAGCAGACTGGAGCAGCTGTACATACCTGGCTGGATTG 85
   ::: ||||| |||
136 yargProSerAlaAspProSerGlyGluLeu.....ValG 148
   .....
86 AGAATGATGAATACATCAATGAATCCACCAAGACCTGAAATTAATGAAGAT 135
   ||||| ||||| ||||| ||||| ||||| |||||
143 luArgArgGluValSerThrLeuGluProLeuProAspValThrAsp 164
   ||||| ||||| ||||| ||||| ||||| |||||
136 GATGTTCAACACAAAGTGTGTTTCAGCAGAAATCCTCGGAGGCTC 185
   ::::: ::::: ::::: |||||
165 ProAlaAspProSerThrIleValGlyAsnArgGluSerValArgLeuAl 181
   ::::: ::::: ::::: |||||
186 ATTTGCTGCTTCT 198
   ||||| ||||| |||||
181 apheValAlaAla 185
   ||||| ||||| |||||
seq_name: pir2:T25190
seq_documentation_block:
hypothetical protein T23G11.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25190
R:Gardner, A.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19993
A:Accession: T25190
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-243 <WIL>
A:Cross-references: EMBL:Z81130; PIDN: CAB03414.1; GSPDB: GN00019; CESP: T23G11.5
A:Experimental source: clone T23G11
C:Genetics:
A:Map position: 1
A:Introns: 41/2; 82/3; 115/3; 199/2
alignment_scores:
Quality: 63.50 Length: 79
Ratio: 1.443 Gaps: 5
Percent Similarity: 55.696 Percent Identity: 27.848
alignment_block:
US-09-049-696-13 x T25190 ..
Align .seg 1/1 to: T25190 from: 1 to: 243
10 CTGGAGGAGTTACGACGCCAGCAGAGAGTATACCCAGCAGATGG 59
   ::::: ||||| ||||| ::::: |||||
32 ValSerGlyHisAsnMetSerTyrAlaArgTyrThrProLeuGlu 48
   .....
63 ACCACTGTACATACCTGCTGCTGATGAGATGATGAATACATGAATC 109
   | ||| ::::: ||||| ||||| ||||| |||||
48 yargLeu...ValGluLysThrLeuGluAspGluLeuLysThrAlaG 64
   .....
110 CACCAAGACCTGAAATTAATGAATGATGATGATGATGATGATGATG 147
   ::::: ||||| ||||| ||||| ||||| |||||
64 InIleHisGluGluLeuThrLysGlnPheLeuLeuAspLeuHis 80
   .....
148 AAGCAAGTGTGTTTCAGCAGAAATCCTCGG..... 178
   ||||| ::::: ||||| ||||| ||||| |||||
81 SerLys...IleSerAlaArgHisGluArgGlyArgThrAlaGlu 96
   .....
179 ..GAGGCTCATTTGCTGCTTCTGATGTCCTCAATG 211
   . ::::: ||||| ||||| ||||| ||||| |||||
96 euAspGluHisAlaTrpGlnValGlnAsnAsnIle 107
   .....
seq_name: pir2:T43522
seq_documentation_block:
condensin complex component cnd3 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe

```


69 GlnValCysPheSerArgThrSerGlyGlySerPheValAlaSerAs 85
201 TGTCCCAATGCTCCACACCTGAT 225
|||||
85 pValProAsnAlaProIleProAsp 93

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US95-07289-9

seq_documentation_block:

; Sequence 9; Application PCT/US9507289
; GENERAL INFORMATION:
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: Colon Specific Genes and Proteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; ADDRESS: Stewart & Olslein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07289
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-265
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-07289-9

alignment_scores:
Quality: 403.00 Length: 75
Ratio: 5.373 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-049-696-13 x PCT-US95-07289-9

Align seg 1/1 to: PCT-US95-07289-9 from: 1 to: 228

1 GTGCGGCTCTGGGAGGTTAAACGACGACGAGAGTGATACCCCA 50
|||||
19 ValArgAlaLeuGlyValAsnAlaArgArgValIleProGI 35
|||||
51 GCAGAGTGGAGCACTGACATACCTGCTGGATGAGATGCAATAC 100
|||||
35 nGlnSerGlyAlaLeuTyrIleProGlyTyrIleGluAsnAspGluIleG 52
|||||
101 AATGGAATCCACCAAGACCTGAAATTAAGGATGATGTTCAACACAAG 150
|||||
52 InTrpAsnProArgProGluIleAsnLysAspValGlnHisLys 68
|||||
151 CAAGTGTGTTTACGACGAACATCTCGGGAGGCTCATTTGTGGCTTCTGA 200
|||||
69 GlnValCysPheSerArgThrSerGlyGlySerPheValAlaSerAs 85

201 TGTCCTCAATGCTCCACACCTGAT 225
|||||
85 pValProAsnAlaProIleProAsp 93

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-462-467B-6

seq_documentation_block:

; Sequence 6; Application US/08462467B
; Patent No. 6210899
; GENERAL INFORMATION:
; APPLICANT: Rosenbaum, Jan S
; TITLE OF INVENTION: The Use of a BMP Protein Receptor
; TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells
; TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I
; TITLE OF INVENTION: BMP Receptor
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 11810 East Miami River Road
; CITY: Ross
; STATE: OH
; COUNTRY: USA
; ZIP: 45061
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,467B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hersko, Bart S.
; REGISTRATION NUMBER: 32,572
; REFERENCE/DOCKET NUMBER: 5474R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 627-0633
; TELEFAX: (513) 627-0260
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-08-462-467B-6

alignment_scores:
Quality: 64.00 Length: 76
Ratio: 1.778 Gaps: 4
Percent Similarity: 47.368 Percent Identity: 28.947

alignment_block:

US-09-049-696-13/rev x US-08-462-467B-6

Align seg 1/1 to: US-08-462-467B-6 from: 1 to: 417

134 TCCTTATTAAATTCAGGTCTTGGTGGATTCCATTGTTATTCATCA..... 90
|||||
222 AlaLeuLeuTyrSerGlyLeuAsnGlyLeuTyrCysTyrSerThrArgpr 238
|||||
89TTCTCAATCCAGCCAGGTATGTACAGTGCTCCA..... 57
|||||
238 oSerGluArgHisIleSerIleLeuGluGlyLeuTyrAlaSerProProA-255
|||||
57 57
255 rgGlyLeuAsnGlyLeuCysTyrSerHisIleSerThrTyrArgGlyLeu 271
|||||
56 ...CTCTGCTGGGTATCATCTCTCGTCTGGCTGCTTAACT..... 18

|||||:|||||
272 GlyLeuCystyrSerValAla.....LeuValAlaLeuThrHisArgTh 286
17CCTCCAGAGCCGC 3
286 rHisArgThrHisArgProArg 295

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-462-467B-10

seq_documentation_block:
; Sequence 10, Application US/08462467B

; Patent No. 6210899

; GENERAL INFORMATION:

; APPLICANT: Rosenbaum, Jan S

; TITLE OF INVENTION: The Use of a BMP Protein Receptor

; TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells

; TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I

; TITLE OF INVENTION: BMP Receptor

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: The Procter & Gamble Company

; STREET: 11810 East Miami River Road

; CITY: Ross

; STATE: OH

; COUNTRY: USA

; ZIP: 45061

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/462,467B

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Hersko, Bart S.

; REGISTRATION NUMBER: 32,572

; REFERENCE/DOCKET NUMBER: 5474R

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (513) 627-0633

; TELEFAX: (513) 627-0260

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 417 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: protein

US-08-462-467B-10

alignment_scores:

Quality: 64.00 Length: 76

Ratio: 1.778 Gaps: 4

Percent Similarity: 47.368 Percent Identity: 28.947

alignment_block:

US-09-049-696-13/rev x US-08-462-467B-10 ..

Align seg 1/1 to: US-08-462-467B-10 from: 1 to: 417

134 TCCTATTATTCAGCTCTTGGTGGATTCCATTGTATTCATCA..... 90

222 AlalaLeuLeuYrSerGlyLeuAsnGlyLeuYrCystyrSerThrArgPr 238

89TTCTCAATCCAGCCAGGTATGTACAGTGCTCCA.... 57

238 oSerGluArgHisIleSerIleLeuGluGlyLeuYrAlaSerProProA 255

57 57

255 rgGlyLeuAsnGlyLeuYrCystyrSerHisIleSerThrYrArgGlyLeu 271

56 ...CTGTGCTGGGTATCACTCTCCGTCTGCTGCTGCTAACT..... 18

272 GlyLeuCystyrSerValAla.....LeuValAlaLeuThrHisArgTh 286

17CCTCCAGAGCCGC 3

286 rHisArgThrHisArgProArg 295

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-462-467B-24

seq_documentation_block:

; Sequence 24, Application US/08462467B

; Patent No. 6210899

; GENERAL INFORMATION:

; APPLICANT: Rosenbaum, Jan S

; TITLE OF INVENTION: The Use of a BMP Protein Receptor

; TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells

; TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I

; TITLE OF INVENTION: BMP Receptor

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: The Procter & Gamble Company

; STREET: 11810 East Miami River Road

; CITY: Ross

; STATE: OH

; COUNTRY: USA

; ZIP: 45061

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/462,467B

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Hersko, Bart S.

; REGISTRATION NUMBER: 32,572

; REFERENCE/DOCKET NUMBER: 5474R

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (513) 627-0633

; TELEFAX: (513) 627-0260

; INFORMATION FOR SEQ ID NO: 24:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 556 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: protein

US-08-462-467B-24

alignment_scores:

Quality: 64.00 Length: 76

Ratio: 1.778 Gaps: 4

Percent Similarity: 47.368 Percent Identity: 28.947

alignment_block:

US-09-049-696-13/rev x US-08-462-467B-24 ..

Align seg 1/1 to: US-08-462-467B-24 from: 1 to: 556

134 TCCTATTATTCAGCTCTTGGTGGATTCCATTGTATTCATCA..... 90

222 AlalaLeuLeuYrSerGlyLeuAsnGlyLeuYrCystyrSerThrArgPr 238

89TTCTCAATCCAGCCAGGTATGTACAGTGCTCCA.... 57

238 oSerGluArgHisIleSerIleLeuGluGlyLeuYrAlaSerProProA 255

57 57


```
238 oSerGluArgHisIleSerIleLeuGluGlyLeuTyrAlaSerProProA 255
57 .....TTCTCAATCCAGCCAGCATGTACAGTGTCCA..... 57
255 rgGlyLeuAsnGlyLeuCysTyrSerHisIleSerThrTyrArgGlyLeu 271
56 ...CTCTGCTGGGTATCACTCTCCGTCCTGCGTAACT..... 18
272 GlyLeuCysTyrSerValAla.....LeuValAlaLeuThrHisArgTh 286
17 .....CCTCCAGAGCCGC 3
286 rHisArgThrHisArgProArgProArg 295
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-462-467B-2

seq_documentation_block:
Sequence 2, Application US/08462467B
Patent No. 6210899

GENERAL INFORMATION:

APPLICANT: Rosenbaum, Jan S
TITLE OF INVENTION: The Use of a BMP Protein Receptor
TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells
TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I
TITLE OF INVENTION: BMP Receptor
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Procter & Gamble Company
STREET: 11810 East Miami River Road
CITY: Ross
STATE: OH
COUNTRY: USA
ZIP: 45061

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462.467B

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Hersko, Bart S.
REGISTRATION NUMBER: 32,572
REFERENCE/DOCKET NUMBER: 5474R
TELECOMMUNICATION INFORMATION:
TELEPHONE: (513) 627-0633
TELEFAX: (513) 627-0260

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 2887 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-08-462-467B-2

alignment_scores:
Quality: 64.00 Length: 76
Ratio: 1.778 Gaps: 4
Percent Similarity: 47.368 Percent Identity: 28.947

alignment_block:

US-09-049-696-13/rev x US-08-462-467B-2 ..

Align seg 1/1 to: US-08-462-467B-2 from: 1 to: 2887

```
134 TCCTTATTATTCAGGCTCTTGTTGGATTCCATTGTATTCATCA..... 90
222 AlaLeuLeuTyrSerGlyLeuAsnGlyLeuTyrCysTyrSerThrArgPr 238
```

```
189 .....TTCTCAATCCAGCCAGCATGTACAGTGTCCA..... 57
238 oSerGluArgHisIleSerIleLeuGluGlyLeuTyrAlaSerProProA 255
57 .....TTCTCAATCCAGCCAGCATGTACAGTGTCCA..... 57
255 rgGlyLeuAsnGlyLeuCysTyrSerHisIleSerThrTyrArgGlyLeu 271
56 ...CTCTGCTGGGTATCACTCTCCGTCCTGCGTAACT..... 18
272 GlyLeuCysTyrSerValAla.....LeuValAlaLeuThrHisArgTh 286
17 .....CCTCCAGAGCCGC 3
286 rHisArgThrHisArgProArgProArg 295
```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-462-467B-8

seq_documentation_block:
Sequence 8, Application US/08462467B
Patent No. 6210899

GENERAL INFORMATION:

APPLICANT: Rosenbaum, Jan S
TITLE OF INVENTION: The Use of a BMP Protein Receptor
TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells
TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I
TITLE OF INVENTION: BMP Receptor
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Procter & Gamble Company
STREET: 11810 East Miami River Road
CITY: Ross
STATE: OH
COUNTRY: USA
ZIP: 45061

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462.467B

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Hersko, Bart S.
REGISTRATION NUMBER: 32,572
REFERENCE/DOCKET NUMBER: 5474R
TELECOMMUNICATION INFORMATION:
TELEPHONE: (513) 627-0633
TELEFAX: (513) 627-0260

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 2887 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-08-462-467B-8

alignment_scores:
Quality: 64.00 Length: 76
Ratio: 1.778 Gaps: 4
Percent Similarity: 47.368 Percent Identity: 28.947

alignment_block:

US-09-049-696-13/rev x US-08-462-467B-8 ..

Align seg 1/1 to: US-08-462-467B-8 from: 1 to: 2887

```
134 TCCTTATTATTCAGGCTCTTGTTGGATTCCATTGTATTCATCA..... 90
222 AlaLeuLeuTyrSerGlyLeuAsnGlyLeuTyrCysTyrSerThrArgPr 238
```

222 AlaLeuLeuTyrSerGlyLeuAsnGlyLeuTyrCysTyrSerThrArgPr 238
:::||||| |||||:::|||||:::||||| |||||
89TTCAATCCAGCCAGGATGATGACAGTGCTCCA... 57
:::||||| |||||:::|||||:::||||| |||||
238 oSerGluArgHisIleSerIleLeuGluGlyLeuTyrAlaSerProA 255
57
255 rgGlyLeuAsnGlyLeuCysTyrSerHisIleSerThrTyrArgGlyLeu 271
56 ...CTCTGGGTGATCACTCGCTGGCTGGCTGCTAACT..... 18
|||||:::|||||:::||||| |||||
272 GlyLeuCysTyrSerValAla.....LeuValAlaLeuThrHisArgTh 286
17CTCCAGAGCCCGC 3
286 rHisArgThrHisArgProArgProA 295
seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-118-270-50

seq_documentation_block:
; Sequence 50, Application US/08118270
; Patent No. 5508384
; GENERAL INFORMATION:
; APPLICANT: Murphy, Randall B.
; APPLICANT: Schuster, David I.
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118,270
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY-2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633

INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 336 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-118-270-50

alignment_scores:
; Quality: 63.00 Length: 37
; Ratio: 2.625 Gaps: 2
Percent Similarity: 64.865 Percent Identity: 35.135
alignment_block:
US-09-049-696-13/rev x US-08-118-270-50

Align seg 1/1 to: US-08-118-270-50 from: 1 to: 336
150 CTGTGTTGAACATCATCTTATTATTTCAGGTCTTGTGGATTCCA... 103
:::||||| |||||:::|||||:::||||| |||||
206 ValValMetAsnIleValValValThrPheAlaIleCysTrpLeuProTy 222
102TTGTATTTCATCATTT.....CTCAATCCAGCCA 75
:::||||| |||||:::|||||:::||||| |||||
222 rHisValTyrPheIleLeuThrAlaIleTyrGlnGlnLeuAsnArgTrpL 239
74. GGTATGTACAG 64
239 ySTyrlleGln 242

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US93-08528-50

seq_documentation_block:
; Sequence 50, Application PC/TUS9308528

; GENERAL INFORMATION:
; APPLICANT: New York University
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08528
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633

INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 336 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US93-08528-50

alignment_scores:
; Quality: 63.00 Length: 37
; Ratio: 2.625 Gaps: 2
Percent Similarity: 64.865 Percent Identity: 35.135
alignment_block:
US-09-049-696-13/rev x PCT-US93-08528-50

Align seg 1/1 to: PCT-US93-08528-50 from: 1 to: 336
150 CTGTGTTGAACATCATCTTATTATTTCAGGTCTTGTGGATTCCA... 103
:::||||| |||||:::|||||:::||||| |||||
206 ValValMetAsnIleValValValThrPheAlaIleCysTrpLeuProTy 222

102TTGATTTCATCATTT.....CTCAATCCAGCCA 75
:|||||: :
222 rHisValtyrPheileLeuThrAlaileTyrGlnGlnLeuAsnArgTrpL 239
74 GGTATGTACAG 64
:|||||: :
239 ystyrlleGln 242

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-162-809-2

seq_documentation_block:
; Sequence 2, Application US/08162809
; Patent No. 5457048
; GENERAL INFORMATION:
; APPLICANT: Pasquale, Elena B.
; APPLICANT: Sajjadi, Ferydoun G.
; TITLE OF INVENTION: NOVEL EPH-RELATED TYROSINE KINASES,
; TITLE OF INVENTION: NOVEL EPH-RELATED TYROSINE KINASES,
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States of America
; ZIP: 92122

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,809
; FILING DATE:

CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 9503
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 951 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-162-809-2

alignment_scores:
; Quality: 63.00 Length: 21
; Ratio: 3.938 Gaps: 0
; Percent Similarity: 76.190 Percent Identity: 47.619

alignment_block:
US-09-049-696-13 x US-08-162-809-2 ..

Align seg 1/1 to: US-08-162-809-2 from: 1 to: 951

97 ATACAATGGAAATCACCAGACCTGAAATTAATAAGGATGATGTTCAACA 146
:|||||: :
288 LeuGluTrpAsnProArgGluThrGlyArgAspValThrTy 304
147 CAACCAAGCTGCT 159
:|||||: :
304 rAsnIleValCys 308

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-374-483-4-
seq_documentation_block:

Sequence 4, Application US/08374483
Patent No. 5880102
GENERAL INFORMATION:

APPLICANT: GEORGE, SAMUEL E.
APPLICANT: BLAZING, MICHAEL A.
TITLE OF INVENTION: ADENOVIRAL VECTOR SYSTEM
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,483
; FILING DATE: 17-JAN-1995

CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 1579-83
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR

INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1015 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-374-483-4

alignment_scores:
; Quality: 62.50 Length: 80
; Ratio: 1.359 Gaps: 5
; Percent Similarity: 57.500 Percent Identity: 26.250

alignment_block:
US-09-049-696-13 x US-08-374-483-4 ..

Align seg 1/1 to: US-08-374-483-4 from: 1 to: 1015

4 CGGGCTCTGGAGGAGTT.....AACGAGCCAGACGGAG 38
:|||||: :
473 LysAlaPheGlyLeuSerAsnGlyMetAlaHisAlaTyrGluCysLe 489

39 AGTGATACCCAGCAGAGTGGAGCACTGTACATACCTGGCTGGATTGAGA 88
:|||||: :
489 uileLeuProArgThrGluGlyProLeuSerLeu.....Trp..... 501

89 ATGATGAATACAAATCGAATCCACCAAGACCTGAAATTAATAAGGATCAT 138
:|||||: :
502 ..LysLysIleSerTrpLysProValAlaProPheProGlnAspGluAsp 517

139 GTTCAACACAGCAAGTGTGTGTTTCAGCAGAACATCTCTCG..... 177
:|||||: :
518 ValArgAlaTyrMetProCysIleIleAspThrAlaThrThrAspAlaG 534

178 .GGAGGCTCATTTTGGCTTCTGTATGTCCTCCCAATGCTCCC 216
:|||||: :
534 yGlyGlyGluTyr.....GlnValProArgSerPro 544

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-374-483-7

```
seq_documentation_block:
; Sequence 7, Application US/08374483
; Patent No. 5880102
; GENERAL INFORMATION:
; APPLICANT: GEORGE, SAMUEL E.
; APPLICANT: BLAZING, MICHAEL A.
; TITLE OF INVENTION: ADENOVIRAL VECTOR SYSTEM
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,483
; FILING DATE: 17-JAN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 1579-83
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1015 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-374-483-7

alignment_scores:
Quality: 62.50 Length: 80
Ratio: 1.359 Gaps: 5
Percent Similarity: 57.500 Percent Identity: 26.250

alignment_block:
US-09-049-696-13 x US-08-374-483-7 ..

Align seg 1/1 to: US-08-374-483-7 from: 1 to: 1015
4 CGGCTCTGGAGGAGT.....AACGCCACGACGAGG 38
:::|||||:::
473 LysAlaPheGlyLeuSerAsnGlyMetAlaHisAlaTyrgluCysLe 489
39 AGTGATACCCAGCAGAGTGGACAGCTGTACATCTGGCTGCATTCACA 88
:::|||||:::
489 uileLeuProArgThrgluGlyProLeuSerLeu.....Trp..... 501
89 ATGATGAATACATGAATGCCACCAAGACCTGAATTATTAAGGATGAT 138
:::|||||:::
502 ..LysLysIleSerTrpPlyProValAlaProPheProGlnAspGluAsp 517
139 GTTCAACACAGCAAGGTGTCTTCAGCAGACATCTCTCG..... 177
:::|||||:::
518 ValArgAlaTyrgMetProCysIleIleAspThrAlaThrThrAspAlaGl 534
178 .GGAGGCTCATTTGCTCTCTGATGTCCCAATGCTCCC 216
|||||:::
534 yGlyGlyGluTyr.....GlnValProArgSerPro 544

seq_name: /cgn2_6/ptdata/2/iaa/6A_COMB.pep:US-08-737-248-12

seq_documentation_block:
; Sequence 12, Application US/08737248
; Patent No. 6114305
; GENERAL INFORMATION:
; APPLICANT: Guemene, Daniel
; APPLICANT: Zadworny, David
; APPLICANT: Karatzas, Costas
; TITLE OF INVENTION: USE OF PROLACTIN FOR PREVENTING AND/OR
; TREATING BIRD BROODINESS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/737,248
; FILING DATE: 28-APR-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR95/00576
; FILING DATE: 03-MAY-1995
; PRIOR APPLICATION NUMBER: FR 94/05550
; FILING DATE: 05-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 989.6411P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 199 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-737-248-12

alignment_scores:
Quality: 60.00 Length: 34
Ratio: 2.609 Gaps: 1
Percent Similarity: 67.647 Percent Identity: 44.118

alignment_block:
US-09-049-696-13/rev x US-08-737-248-12 ..

Align seg 1/1 to: US-08-737-248-12 from: 1 to: 199
156 CACTTGCTGTGTGAACATCATCTTATTATTTCAGGCTCTTGGTGGAT 107
||| :::|||||:::
77 HisGluValLeuLeuAsnLeuIleLeuArgValLeuArgSerTipAsnAs 93
106 TCATTGTATTTCATCTCAATCCAGCCAGGTATGTACAGTCTCCA 57
|||||:::
93 pProLeuTyHisLeuValThrGluValArg.GlyMetGlnGluAlaPro 109
```


**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER: _____**

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)
